2016 CAS Exam 9 Study Manual

Richard Goldfarb, FCAS, CFA

Version: November 6, 2015

Introduction

Overview

These notes have been prepared to emphasize and explain the material expected to be tested on the CAS exam, based on the published Learning Objectives and Knowledge Statements. In preparing these notes, I set out to achieve the following main goals:

1. Highlight the most important elements of each reading

2. Explain the more difficult material using different language and, in most cases, less mathematics than the required readings

3. Provide numerous sample questions specifically designed to reinforce your understanding of the most important and most difficult concepts, along with detailed explanatory solutions

Because a deep understanding of the concepts is the most effective aid in remembering the details needed for the exam, these notes are not simply outlines of the readings and are not intended to be a source of details to be memorized. Instead, I have carefully selected the most important material from the syllabus and attempted to explain it fully in my own words. For the easier readings, my notes will be rather limited. In these instances you may want to just focus on the sample questions and solutions.

In addition, I recommend also reading the source material in the syllabus. Some of the content found in the source material may have been excluded from these notes if it did not appear to be included within the Learning Objectives and Knowledge Statements.

Condensed Notes for Each Reading

Once you have mastered the material for each reading, you can refer to a separate document on my website entitled *Reading Highlights: Key Concepts and Formulas*, which contains condensed summaries of the notes for each reading.

Recommended Order of Study

The order of the various readings in these notes is consistent with the syllabus. I recommend following the order of the readings as they appear in these notes.

Sample Questions

I have included hundreds of sample questions that are specifically designed to help you focus on the most important concepts and the most difficult numerical problems. Questions that are critical to understanding the important concepts from each reading appear at the end of the reading.

At the end of each section, I include my own solutions to **selected** questions from old CAS exams. See the comments below to understand why I have included only some of the past exam questions in these notes.

The detailed solutions provided for all of the sample questions are intended to guide you through the thought processes required to solve the problems. In many cases the solutions provide further explanations and additional insights not emphasized in the reading summaries. Keep in mind that these solutions are intended to *teach* the material; they are not intended as recommendations for how to answer questions on the exam. In most cases your exam solutions can and should be less detailed than the solutions shown here.

Textbook Questions

The BKM and Hull readings have extensive end-of-chapter questions, many of which have been used verbatim, or nearly so, on past exams. Solutions manuals are available for both of these texts (through Amazon and/or at most college bookstores), so these textbook questions are highly recommended.

For the BKM textbook, I strongly encourage students to work through *all* of the end of chapter problems, especially those marked as "CFA Questions" or "Kaplan Schweser" in the 10^{th} Edition.

For the Hull textbook, you should work through all of the exercises presented within the chapters, but not all of the end of chapter problems are particularly relevant for the exam. I provide recommended practice problems from the end of chapter problem sets for each of the Hull chapters. These recommendations appear right after the practice problems at the end of each chapter in these notes.

Mastery of these questions should be helpful for the exam. Just be careful to focus primarily on questions pertaining directly to the published learning objectives. And pay particular attention to the fact that in many cases important learning objectives related to the BKM material may not be adequately covered in the textbook's end of chapter questions.

Additional Practice Questions - Organized by Section

In addition to the questions that appear after each reading, I have also prepared close to 200 additional questions covering everything that I think is critical for the exam and organized these by section, corresponding to the seven sections of the study manual. These questions can be found on the downloads page of the website, under the file name *Practice Questions by Section (with Solutions).* I recommend that you work through these questions as you complete each section of the study manual. However, you may also decide to use these questions as a section-by-section review after you've completed the entire study manual.

There are two points to note about these Additional Practice Questions:

1. To ensure that you see a LOT of problems, I have tried to minimize the amount of numerical calculations in these practice questions. Actual exam questions tend to require a lot of calculations, but these are time consuming and (in my opinion) not the ideal way to spend your time in advance of the exam. It's better to spend time covering more questions that emphasize the underlying concepts and methods than to spend time doing 15 minutes of algebra for a single question.

2. The practice questions were specifically designed to test at higher levels on Bloom's Taxonomy, in accordance with the most recent efforts by the CAS to alter the style of exam questions. These questions will appear *different* from the older CAS exam questions, which tended to ask direct calculation or definitional questions, but should be more consistent with the style of questions the CAS is beginning to use.

Practice Exams

Some students may prefer to have the Additional Practice Questions organized into comprehensive practice exams and use them in the final stages of preparation for the exam. The nearly 200 practice questions are therefore also available in this format as a series of seven comprehensive practice exams. They are the same set of questions, just organized differently. You can download the practice exams from the website either with or without the solutions shown, in files labeled *Comprehensive Practice Exams*.

Comments Regarding Past CAS Exams

At the end of each major section of these notes, I have included tables listing all past CAS exam questions by reading and have included my own sample solutions for *some* of these past CAS exam questions. The questions for which I include solutions were selected either because I find them to be particularly important for the current exam, because they cover a topic that my own questions don't specifically address or because I don't think that the published solution is sufficient for your exam preparation.

In many cases exam questions that are similar or identical to the questions contained in my own end-of-reading questions (even those that have appeared repeatedly on past exams) may not be included here. In your own preparation, you should indeed work through as many variations on the same problems as possible, as subtle differences could trip you up on the actual exam.

For those students who want to study additional old exam questions other than those included here, please keep in mind two points.

First, all of the readings in the *Rate of Return and Risk Load* section of the syllabus have been on the old Exam 9 syllabus for a number of years and were a significant portion of that exam. As a result, there are a large number of old questions from those exams. By and large though these questions are simple applications of the material in the readings and are not materially different from the practice questions contained in these notes or the numerical examples in the readings. I have included several of these in the notes from recent exams, but you may want to review others going back further to see minor variations in wording.

Second, most of the other readings on the syllabus have been on the old Exam 8 syllabus for a number of years. In the case of the BKM readings, old exam questions back to 2000 can be referenced. I have included questions from the old Exam 8 exams that I think are useful. I don't recommend spending a lot of time on too many of the others that I didn't include.

Study Manual Updates

While there are no specific changes planned, there may be edits, revisions or new material added to the manual prior to the exam date. All updates will be incorporated into the manual and a new version of the manual will be posted, in its entirety, to the downloads section of the website. Any changes that are made after the initial release of the manual will be listed, along with the date the change was made and the affected page number(s), in the Appendix.

I will send announcements of *material* changes to the study manual via email to all students registered on my website, but please be sure to check the website from time to time for minor updates or other announcements.

About the Author

Richard Goldfarb, FCAS, CFA

Richard is the Group Chief Financial Officer and Chief Actuary of CapSpecialty, a specialty insurance subsidiary of Alleghany Corporation. In his role at CapSpecialty, Richard is responsible for financial reporting, capital management, financial risk management, loss reserving and pricing. Over his 26 year career, Richard has held a variety of financial and actuarial roles within the insurance industry, giving him a broad range of experience in financial risk management, capital management and reinsurance.

Richard is a Fellow of the Casualty Actuarial Society, received his MBA in Finance from the University of Chicago, and is a CFA charterholder.

Richard was formerly a member of the Board of Directors of the CAS and has previously been a member of the CAS Valuation, Finance and Investment Committee; the CAS Enterprise Risk Management Committee; the SOA Task Force on ERM Credentials, and a consultant to the CAS Exam 8 Syllabus Committee. He is the author of two readings on the CAS syllabus and has been a frequent presenter at industry conferences on topics including enterprise risk management, economic capital modeling and reinsurance pricing.

Since 1993, Richard has taught actuarial exam seminars and written study manuals for several CAS and SOA actuarial exams, including CAS Exam 5B (Finance), CAS Exam 8 (Investments), SOA Course 8V (Investments), the SOA Financial Economic Theory Exam and the SOA Advanced Portfolio Management Exam.

Part 1

Portfolio Theory

Part 1 Table of Contents

BKM Chapter 6: Risk and Risk Aversion

Risk Aversion

One of the fundamental questions in financial theory is whether a given opportunity offers a sufficient risk premium, defined as the expected return in excess of the risk-free rate, to compensate an investor for the risk that he is exposed to. We will generally assume that investors are **risk averse** in the sense that they would not accept a risky proposition without some degree of compensation in the form of a risk premium. The question then becomes, "How much of a risk premium is sufficient?"

Utility

In order to rank the appeal of various risky prospects, it is useful to be able to quantify investors' risk aversion. This is done using a **utility score** to determine if one risky prospect is better than another (see the notes for Appendix 6A for a more complete discussion of why it is important to evaluate the utility of payoffs rather than the dollar value of payoffs). While such a utility measure is primarily a theoretical value, we can assume that it is related to the expected return and variance of returns (a standard measure of risk). That is to say, for a given expected return the utility will be lower as the variance of returns increases.

There are an infinite number of utility functions we could assume, but for our purposes we will assume that utility can be measured as:

$$U = E(r) - .5A\sigma^2$$

with A representing the investor's degree of risk aversion (a higher value means they assign a larger penalty for risk).

Note that in the previous versions of the textbook (and therefore on older exam questions) the "0.5" factor was "0.005" and the expected return and standard deviation parameters were entered using whole numbers rather than decimals (e.g. they used to use "4" rather than ".04" to represent 4%).

This framework then allows us to distinguish between the following types of investors:

- **Risk Averse Investors**: Their risk aversion parameter, A, is greater than zero. Their utility score is adjusted downward when there is risk.
- **Risk Neutral Investors**: Their risk aversion parameter, A, is equal to zero. They judge prospects solely in terms of their expected return and do not modify their utility score when there is risk.
- **Risk Seeking Investors**: Their risk aversion parameter, A, is less than zero. Their utility score is adjusted *upward* when there is risk (they actually prefer risky investments to risk-free assets).

Mean-Variance Criterion

Once we can quantify an investor's attitude about risk, we can rank any two opportunities for that investor. To do so, we note that portfolio A would be preferred to portfolio B if $E(r_A) \geq E(r_B)$ and $\sigma_A \leq \sigma_B$. That is, the portfolio is preferred if its expected return is at least the same or higher and its risk is at least the same or lower.

Certainty Equivalent Rate

With the particular utility function we are assuming here, we could interpret the utility value as the rate of return that would cause the investor to be indifferent between the risky investment and the risk-free investment. They will therefore accept the risky investment if the certainty equivalent rate is above the risk-free rate and will reject it if it falls below the risk-free rate.

Indifference Curves

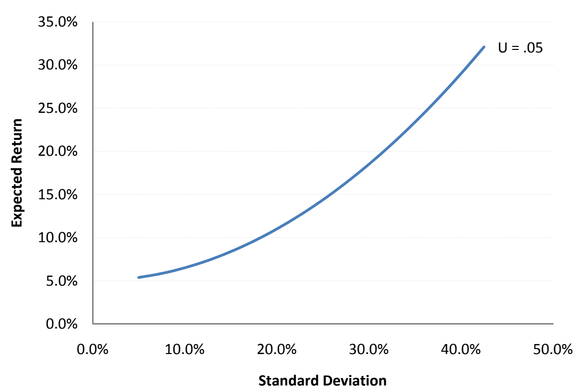
It will be helpful to be able to identify all of the opportunities that the investor will be *indifferent* to, meaning that all of them have the same utility score. This collection of opportunities can be plotted as an *indifference curve*, with each point on the curve representing a combination of risk and return that provides the same level of utility.

As a simple example, consider an investor with risk aversion $A = 3$ and utility specified as:

$$U = E(r) - .5A\sigma^2$$

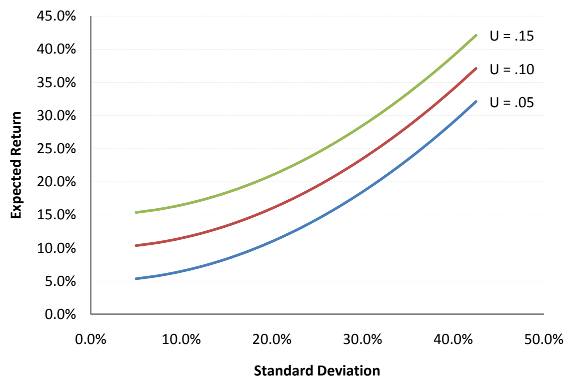
The various combinations of expected return and standard deviation which all result in a utility score of .05 are shown in Figure 1.

FIGURE 1. Sample Utility Curve $U = .05$



Similarly, Figure 2 on the next page shows the combinations that result in a utility score of .10 or .15 as well.

FIGURE 2. Comparison of Utility Curves



Estimating Risk Aversion

An investor's risk aversion parameter, A, cannot be directly observed. However, it may be possible to infer it based on what they are willing to pay to insure against a loss. For instance, assume that an investor would be *indifferent* between taking on a risk with a given expected return and standard deviation or paying v for an insurance policy. The insurance policy, along with the original risk, results in a portfolio with an expected return of $-v$ and a standard deviation of zero.

Their indifference between these two possibilities suggests that they have the same utility score, so we can set the utility scores equal to each other and then solve for A.

See the numerical problems below for examples of this calculation.

Capital Allocation Across Risky and Risk-Free Portfolios

The most important aspect of this chapter is the discussion of the expected return and standard deviation (risk) combinations that result from various combinations of a risky portfolio and a risk-free portfolio.

We will assume that there exists just one risky portfolio, with an expected return and a non-zero standard deviation, and one risk-free portfolio with an expected return and zero standard deviation. The investor then simply must choose how much weight to give each of these two portfolios. Each choice will result in a different combination of risk and expected return.

When these results are combined with the utility concepts above, we can begin to understand how an investor would choose an optimal portfolio from among the wide range of choices for different weights, y, assigned to the risky portfolio.

- *Risky Portfolio* — We assume that the risky portfolio is some pre-determined combination of risky assets, which could include stocks, long-term bonds, real estate, art, etc. The expected return is denoted $E(r_p)$ and the standard deviation is denoted σ_p.

- *Risk-Free Portfolio* — This is generally assumed to be a US Treasury bill because it is free of default risk, the short-term nature makes it mostly insensitive to interest rate risk and its exposure to inflation uncertainty is minimal. But in practice, all money market instruments (including CDs and Commercial Paper) can be viewed as essentially risk-free in this sense. The return on the risk-free portfolio is denoted r_f.

- *Combinations of Risky and Risk-Free Portfolios* — The two key values we care about are the expected return and standard deviation of a *complete* portfolio, c, that consists of y percent of the total portfolio value invested in the risky portfolio, p, and $(1 - y)$ invested in the risk-free portfolio.

Expected Return of Complete Portfolio

The following formula shows the expected return on the complete portfolio:

$$\begin{aligned} E(r_c) &= yE(r_p) + (1 - y)r_f \\ &= r_f + y[E(r_p) - r_f] \end{aligned}$$

Notice that by rewriting the formula the second way, we can see that the return on our complete portfolio is equal to the risk-free rate plus a *risk premium*. The risk premium for the complete portfolio is the product of the risk premium on the risky portfolio equal to $[E(r_p) - r_f]$ and the percentage invested in the risky portfolio, y.

Std. Deviation of Complete Portfolio

As we will see in the next chapter, the variance of a combination of any two assets (denoted generically as asset 1 and asset 2) is given by the following equation:

$$\sigma^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2Cov(r_1, r_2)$$

Since the standard deviation of the risk-free asset is zero and its covariance with the risky portfolio is also zero, the variance of the complete portfolio when the weight in the risky portfolio is y can be written as follows:

$$\sigma_c^2 = y^2\sigma_p^2$$

Or, after taking the square root of both sides this can be written as:

$$\sigma_c = y\sigma_p$$

Capital Allocation Line

If we plot combinations of risk (as the x-axis) and expected return (as the y-axis) for our combination of the risk-free asset and the risky asset based on the above two formulas, we produce what is known as the **Capital Allocation Line** (CAL).

- One point on this line is the intercept, when $y = 0$ then $E(r_c) = r_f$ and $\sigma_c = 0$.
- Another point on this line is when $y = 1$ and therefore $E(r_c) = E(r_p)$ and $\sigma_c = \sigma_p$.

Using these two points, we can calculate the slope of the line that connects the two points as:

$$S = \frac{E(r_p) - r_f}{\sigma_p}$$

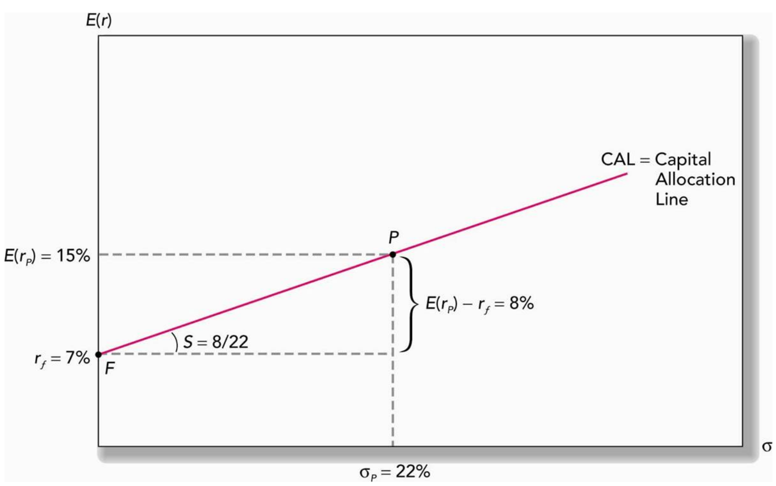
Note that this slope is called the reward-to-variability ratio or the Sharpe ratio.

Given the slope and the intercept, we can write the equation for the relationship between the expected return and the standard deviation as:

$$E(r_c) = r_f + \frac{\sigma_c}{\sigma_p}[E(r_p) - r_f]$$

The textbook's Figure 6.4, shown here as Figure 3, shows the CAL with some values inserted.

FIGURE 3. Capital Allocation Line



Leverage and Borrowing vs. Lending

Notice that we never imposed any restriction on y, the percentage invested in the risky portfolio. If $y < 1$ then some of our money is invested in the risk-free asset and some in the risky portfolio. We can think of the amount invested in the risk-free asset as lending money (to the US government in the case of US T-Bills).

But if $y > 1$, then more than our initial portfolio value is invested in the risky asset, which must mean that we are borrowing some amount in order to do this. This of course is the same as saying that we are investing a negative amount in the risk-free asset.

Risk Tolerance and Asset Allocation

So far, we have simply shown the combinations of risk and return that are possible. To determine which particular portfolio someone would want to own, we need to incorporate the investor's attitudes about risk. For instance, someone who doesn't want to take any risk would just invest 100% in the risk-free asset. Alternatively, someone who wanted to take more risk than can be achieved than investing in the risky asset would borrow at the risk-free rate and invest more than 100% of their initial portfolio value in the risky asset.

So how is the specific portfolio selected? Recalling the concept of utility, the investor would simply select the portfolio that maximizes his or her utility. To do this, we write the formula for the investor's utility function and select the proportion in the risky asset, y, that maximizes this function.

If $U = E(r) - .5A\sigma^2$ (the typical utility function assumed in the text) then we can just plug in our formulas for expected return and variance of our complete portfolio to get:

$$U = r_f + y[E(r_p) - r_f] - .5Ay^2\sigma_p^2$$

Taking the derivative with respect to y and setting the derivative equal to zero we can solve for the optimal y to get:

$$y^* = \frac{[E(r_p) - r_f]}{A\sigma_p^2}$$

Graphical Interpretation

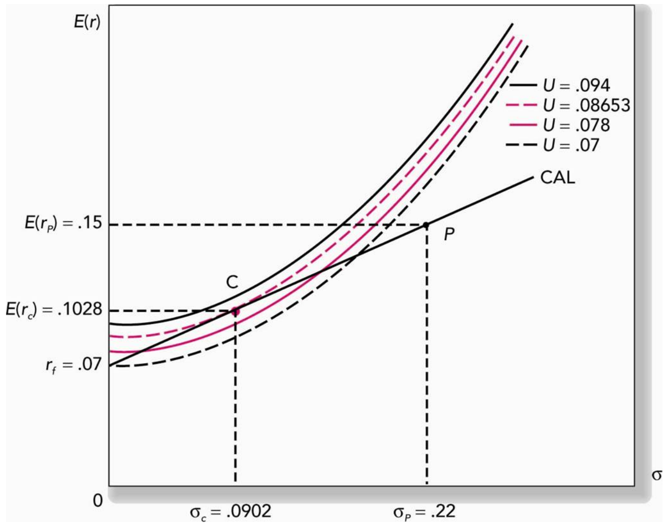
Referring back to the concept of the indifference curves, we could find the optimal complete portfolio by graphing the CAL and then searching for the indifference curves representing the highest utility level that is tangent to the CAL.

For example, assuming $U = E(r) - .5A\sigma^2$ and $A = 4$, $r_f = 7\%$, $E(r_p) = 15\%$ and $\sigma_p = 22\%$, the textbook's Figure 6.8, shown here as Figure 4 on the next page, shows the optimal complete portfolio as the point that is tangent to the indifference curve with $U = 0.08653$.

Notice in the graph that the selected optimal portfolio for this investor, depicted by the tangency point C, has an expected return of 10.28% and a standard deviation of approximately 9.02%. We can confirm this result using the formula shown earlier to identify the optimal portion of the portfolio to invest in the risky portfolio:

$$\begin{aligned} y^* &= \frac{[E(r_p) - r_f]}{A\sigma_p^2} \\ &= \frac{[.15 - .07]}{4(.22^2)} \\ &= 0.41 \end{aligned}$$

FIGURE 4. Optimal Complete Portfolio



From this, we can then find the expected return (using the rounded value of y^*, to two decimal places) as:

$$\begin{aligned} E(r_c) &= r_f + y^*[E(r_p) - r_f] \\ &= .07 + .41[.15 - .07] \\ &= 10.28\% \end{aligned}$$

Passive Strategies and the Capital Market Line

We began the chapter by stating that the risky portfolio could be constructed in many ways and left the details for a later chapter. But assume that instead of spending time and resources to identify any particular risky portfolio we simply chose as the risky portfolio a large, well diversified portfolio of stocks representing the entire US corporate sector? One such portfolio might be the S&P 500, as discussed in Chapter 2 of BKM.

If we used T-bills as our risk-free asset and a broad index of common stocks such as the S&P 500 as our risky asset, then the Capital Allocation Line we developed here could be referred to as the **Capital Market Line** (CML).

The main body of the text provides two main reasons why this passive strategy may make sense — it minimizes costs to acquire the information necessary to select an optimal risky portfolio and it takes advantage of everyone else's efforts to do so, resulting in security prices which are by and large fair.

The side bar also addresses common criticisms of the passive strategy (known as indexing):

- They're Undiversified — Although the S&P 500 holds 500 stocks, it weights them according to their market capitalization and therefore 25% is invested in the top ten firms. However, actively managed funds are similarly undiversified, with 36% on average invested in their top ten holdings.

- They're Top Heavy — Although the S&P 500 only includes 500 of the more than 6,000 US stocks, this does represent 77% of the US market's total market value.

- They're Chasing Performance — As a stock's value increases, the S&P 500 index funds own more of it. But this is not unique to index funds. This always happens because when a stock's market value rises, collectively all shareholders own more of it. This real issue is whether investors wind up owning proportionately more of the stocks whose price rises *above its true value.*

- You Can Do Better — While you may get lucky from time to time, on average investors cannot beat a market index because collectively all investors are the market. Once you factor in the substantial difference in transaction and research costs, the evidence shows that few active fund managers beat the S&P 500 or broader indices.

Practice Questions

Question 1. Suppose that an investor's utility curve can be characterized by the function $U = E(r) - .5A\sigma^2$ with $A = 4$. What would be the investor's utility value for an investment with an expected return of 16% and a standard deviation of returns equal to 20%?

Solution. The investor's utility would equal $U = .16 - .5(4)(.20^2) = .08$.

Question 2. Suppose that the risk-free rate of return were 7%. Would the investor in the previous question prefer to invest in the risk-free asset or in a risky asset with expected return equal to 18% and a standard deviation of 20%?

Solution. Here, the utility value for the risky asset would equal $U = .18 - .5(4)(.20^2) = .10$. Since this exceeds the rate of return on the risk-free asset (note that the utility of the risk-free asset is $U = .07 - .5(4)(0) = .07$), the risky asset is preferred.

Question 3. What would the risk-free rate of return need to be in the previous question in order for the risk-free asset to be preferred to the risky asset?

Solution. The question is asking for the certainty equivalent rate, which is 10%. The risky asset with an 18% expected return is as attractive as a risk-free asset with an expected return of 10% and so unless the risk-free rate exceeded 10%, the risky asset would be preferred.

Question 4. Suppose that you have two risky assets, one with an expected return of 20% and a standard deviation of 25% and the other with an unknown expected return and a standard deviation of 30%. If we knew that an investor with an index of risk aversion $A = 5$ was indifferent between these two risky assets, what is the return on the second risky asset? Assume that the investor's utility curve is given by the equation $U = E(r) - .5A\sigma^2$.

Solution. The fact that the investor is indifferent between these two assets tells us that the utility values are identical. In the case of the first asset, would equal:

$$U = .20 - .5(5)(.25^2) = .0438$$

Setting the utility of the second asset equal to 4.38% and solving for $E(r)$ gives us:

$$\begin{aligned} E(r) &= U + .5A\sigma^2 \\ &= 4.38\% + .5(5)(.30^2) \\ &= 26.88\% \end{aligned}$$

Question 5. Suppose your utility function were given by the equation $U = E(r) - .5A\sigma^2$ with risk aversion coefficient $A = 2$. Your initial wealth is \$500,000, all of which reflects the value of your home. There is only a 1% probability of a hurricane hitting your home, but if it does you will suffer a total loss. The expected loss, expressed as a percentage of your initial wealth, is therefore $p(-100\%) + (1 - p)(0\%) = -.01$. The variance of your losses can be found as the expected squared deviation from the mean, which for such a binomial distribution is simply $p(1 - p) = .0099$. Determine what premium you would pay for an insurance policy such that your utility is the same with or without the insurance policy. Assume that in the absence of a hurricane loss your total wealth will be unchanged at the end of the period.

Solution. Without the insurance policy, your utility is $U = E(r) - .5A\sigma^2 = -.01 - .5(2)(.0099) = -.0199$. With the insurance policy, your return is guaranteed to be 0% with no standard deviation, but for the price of the insurance policy. If the premium paid is v (as a percentage of the value of the house), then your utility is simply $-v$.

Setting these equal, we see that the fair premium is such that $-v = -0.0199$ or $v = 0.0199$.

Question 6. Continuing with the previous question, what multiple of the expected loss were you willing to pay for the insurance policy? How would this multiple have changed if your risk aversion coefficient were $A = 4$?

Solution. The expected loss was .01 and we were willing to pay .0199. This is a multiple of $m = .0199/.01 = 1.99$.

Had we been even more risk averse, with $A = 4$, the premium we were willing to pay would have been 0.0298, which would have been a multiple of 2.98.

Question 7. Suppose you wanted to create a portfolio consisting of 25% invested in a risky portfolio with an expected return of 15% and a standard deviation of 20% and 75% invested in the risk-free asset with expected return equal to 5%. What would be the expected return and standard deviation of the complete portfolio?

Solution. Here, we know that the expected return of the portfolio is given by:

$$\begin{aligned} E(r_c) &= yE(r_p) + (1-y)r_f \\ &= r_f + y[E(r_p) - r_f] \\ &= 5 + .25[15-5] \\ &= 7.5 \end{aligned}$$

and the standard deviation is given by:

$$\begin{aligned} \sigma_c &= y\sigma_p \\ &= .25(20) \\ &= 5 \end{aligned}$$

Question 8. Assume an investor with \$1,000 wanted to borrow \$500 at the risk-free rate and invest all of his wealth in the same risky asset as in the previous question. What would be this investor's complete portfolio return and standard deviation?

Solution. Here, the investor is using borrowed funds to leverage his investment, which should result in higher expected return but higher risk as well. In terms of the variables used, this means that $y > 1$. In fact, since the total investment in the risky asset will be \$1,500 and the initial wealth was only \$1,000, $y = 1{,}500/1{,}000 = 150\%$.

Therefore,

$$\begin{aligned} E(r_c) &= yE(r_p) + (1-y)r_f \\ &= r_f + y[E(r_p) - r_f] \\ &= 5 + 1.5[15-5] \\ &= 20\% \end{aligned}$$

and

$$\sigma_c = y\sigma_p = 1.5(20) = 30\%$$

Question 9. Assume that an investor has a utility function of the form $U = E(r) - .5A\sigma^2$ and a coefficient of risk aversion $A = 6$. What proportion y invested in the risky asset would result in the maximum possible utility?

Solution. To maximize the utility, we simply take the derivative of the utility function with respect to y and set it equal to zero. In order to write the utility function in terms of y though, we need to first plug in the formulas for $E(r)$ and σ in terms of y into our formula for utility. In symbols,

$$\begin{aligned} U &= E(r) - .5A\sigma^2 \\ &= r_f + y[E(r_p) - r_f] - .5A(y\sigma_p)^2 \end{aligned}$$

The derivative is then equal to:

$$\frac{dU}{dy} = [E(r_p) - r_f] - 2(.5)Ay\sigma_p^2 = 0$$

which can be solved for the optimal allocation to the risky portfolio as:

$$\begin{aligned} y^* &= \frac{[E(r_p) - r_f]}{A\sigma_p^2} \\ &= \frac{.15 - .05}{(6).20^2} \\ &= .417. \end{aligned}$$

Important Note — The textbook is somewhat sloppy in that they sometimes use 20 to indicate the 20% standard deviation and they sometimes use .20. Whenever you are using utility function calculations, or more specifically whenever the risk aversion parameter A is present, you should use the decimal point version otherwise your units will be off. In many cases it won't matter which approach you use, which explains the inconsistent treatment in the text, but for the exam you should be careful.

Question 10. What is the expected return and standard deviation of the optimal portfolio found in the previous question? How does the utility score for this portfolio compare to the combinations in the previous questions with either 25% invested in the risky asset or with \$500 of additional borrowing? Assume the investor's risk aversion coefficient is $A = 6$.

Solution. Plugging in the optimal $y = .417$ into the same formulas as before, we find that $E(r) = 9.17\%$ and $\sigma = 8.34\%$.

This produces a utility score of $U = .0917 - .5(6)(.0834^2) = .07083$.

Note that this utility is higher than what we would get from the combinations containing either 25% in the risky asset ($U = 0.0675$) or 150% in the risky asset ($U = -0.07$).

Note that if other values of y were tested, the utility values would be as follows, demonstrating that $y = .417$ does achieve the highest utility value.

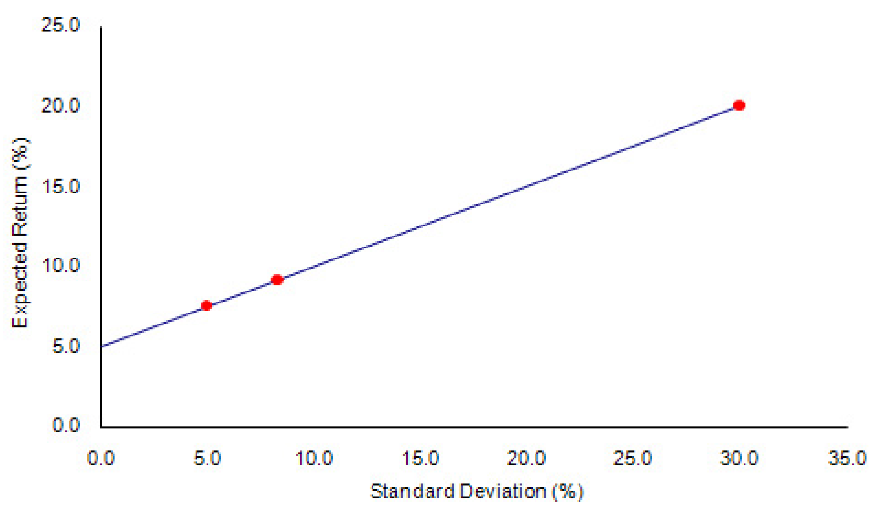
TABLE 1. Comparison of Utility — Different Values of y

y	Utility
0.000	0.05000
0.100	0.05880
0.200	0.06520
0.300	0.06920
0.400	0.07080
0.417	**0.07083**
0.500	0.07000
0.600	0.06680
0.700	0.06120
0.800	0.05320
0.900	0.04280

Question 11. Graph the Capital Allocation Line for the risky asset used in the previous questions using the expected return and standard deviations of the portfolios with risky asset weights $y = 25\%$, $y = 41.7\%$ and $y = 150\%$.

Solution. Using the results from the previous questions, the graph is as follows:

FIGURE 5. Capital Allocation Line — Different Values for y



Question 12. What three characteristics do U.S. Treasury Bills have that make them an appropriate choice as the risk-free asset?

Solution. They are free of default risk, their short term nature makes them insensitive to interest rate risk and their exposure to inflation uncertainty is minimal.

Question 13. Assume that 75% of all invested assets in the US are invested in risky assets and 25% are invested in the risk-free asset. Assume the economy-wide risky asset portfolio has the same risk premium as the historical risk premium on the S&P 500 equal to 8.4% and a standard deviation of 20.5%. Further assume that investors' utility can be depicted as $U = E(r) - .5A\sigma^2$ and that they select their optimal portfolio mix by maximizing their utility. Use this to estimate the average coefficient of risk aversion, A.

Solution. Since US investors follow a utility maximization strategy, then we know that:

$$\begin{aligned} y^* &= \frac{E(r_p) - r_f}{A\sigma_p^2} \\ &= \frac{.084}{A(.205^2)} \\ &= .75 \end{aligned}$$

Using this, we can solve for $A = 2.7$.

BKM Appendix 6A
Risk Aversion, Expected Utility and the St. Petersburg Paradox

St. Petersburg Paradox

Consider the following gamble:

> You pay \$$X$ to play the game and then you flip a fair coin until a head appears. The number of tails that appeared before the first head appeared, n, determines the payoff from the game, Payoff = \$$2^n$. For example, if a head appears on the fourth toss you get \$$2^3$ = \$8.

How much would you pay to play this game?

If you calculated the expected payoff from the game, you would find that the expected value is infinite. But surely nobody would pay an infinite amount of money to play the game. Hence the paradox—the game has an infinite expected payoff and yet there's a finite entry fee that someone would pay to play.

To resolve the paradox, one needs to assume that you would evaluate the payoffs not in terms of dollars but in terms of units of utility. Of course, this would not necessarily resolve the paradox unless the utility value exhibited the property of decreasing marginal utility for each successive dollar of payoff. For instance, the eighth dollar is less valuable in utility units to the person than the seventh dollar was. To put it in more mathematical terms, the utility function must be concave.

Declining Marginal Utility of Wealth

There are an infinite number of concave utility functions that could be assumed. More importantly, every investor could conceivably have a different utility function—for instance because they begin with dramatically different levels of wealth. However, in both cases, the per-unit increment in utility decreases as wealth increases.

Certainty Equivalent Value

This is the riskless wealth level that provides the same expected utility as a combination of your initial wealth plus the outcome of a risky gamble. The difference between the expected wealth level and the certainty equivalent amount reflects the penalty that risk averse people place on the risky outcomes to compensate them for the risk.

Practice Questions for Appendix

Question 14. Suppose that you have wealth equal to \$50,000 and you were given the prospect of a gamble with a 30% chance of winning \$30,000 and a 70% chance of losing \$10,000. Assuming that your utility function equaled $ln(Wealth)$, what is the expected utility of this gamble? Would you accept this gamble?

Solution. If you accept the gamble, your wealth will either be \$80,000 or \$40,000. From this, we can determine the Expected Utility is equal to:

$$E(U) = (.30)ln(80{,}000) + (.70)ln(40{,}000) = 10.804$$

Notice that with beginning wealth of \$50,000, the utility is already equal to $ln(50{,}000) = 10.82$. Therefore, this gamble is expected to result in a lower utility than not taking the gamble and therefore you would not take the gamble, even though the expected gain in dollar terms is $.3(30{,}000) - .7(10{,}000) = 2{,}000$.

Question 15. What is the certainty equivalent value of the gamble presented in the previous question?

Solution. The certainty equivalent of this gamble represents the fixed dollar amount with the same expected utility as the gamble, so $10.804 = ln(W_{CE})$ and solving for $W_{CE} = 49{,}245.78$.

Question 16. Does the utility function presented above reflect risk aversion?

Solution. Notice that with the gamble, the expected wealth is $.3(80{,}000) + .7(40{,}000) = 52{,}000$. However, this has an expected utility of only 10.804.

This same utility could be achieved with a riskless wealth of 49,245.78, the certainty equivalent of the gamble.

Since the certainty equivalent is lower than the expected wealth, we say that this investor is *risk averse.* Although the expected wealth is 52,000, they assign a penalty equal to 2,754.20 to this because of the uncertainty.

BKM Chapter 7: Optimal Risky Portfolios

This chapter extends the results from the previous chapter to the case where we have to first determine the composition of the risky asset portfolio and then determine the Capital Allocation Line depicting combinations of this risky portfolio with the risk-free asset. Then, as before, investors will select a particular portfolio of risky and risk-free assets based on their specific risk tolerance.

Diversification

There are two broad sources of uncertainty in the returns on risky assets—those related to general economic conditions and those which are firm specific. When multiple risky assets are combined into a single portfolio, the firm specific risks will tend to offset each other (to various degrees). The result, referred to as diversification, is that the total firm specific risk will be less than the sum of the individual stocks' firm specific risks.

Notice though that certain risks related to general economic conditions will be common to all risky assets and therefore these systematic risks will not be reduced through diversification.

Portfolios of Two Risky Assets

To put the above point in a more mathematical context, consider the risk and return of portfolios of risky assets. If we invest proportions w_1 and w_2 in two risky assets, the expected return is:

$$E(r_p) = w_1 E(r_1) + w_2 E(r_2)$$

The variance of the return, which we use to reflect the portfolio risk, is:

$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 Cov(r_1, r_2)$$

Notice that we can write the covariance term as a function of the standard deviations and the correlation between the two risky assets:

$$Cov(r_1, r_2) = \sigma_{12} = \sigma_1 \sigma_2 \rho$$

Then the portfolio variance can be written as:

$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \sigma_1 \sigma_2 \rho$$

This form will be convenient in some instances.

From these two equations for the expected return and risk of a portfolio of two risky assets we can see that the expected return for a portfolio is always the weighted average of the expected return of its components. However, the variance — and hence the standard deviation — will depend on the variances as well as the correlation of the assets.

Risk and Return Trade-Offs

Using the previous equations and varying the amount invested in assets 1 and 2, referred to as Stock 1 and Stock 2, we can graph specific combinations of expected returns and standard deviations for portfolios with different weights on two assets (refer to the questions below for the numbers behind these graphs). In this example, shown in Figure 1, I assume that the correlation between the two assets is $\rho = .453$.

FIGURE 1. Risk and Return — Two Assets with $\rho = 0.453$

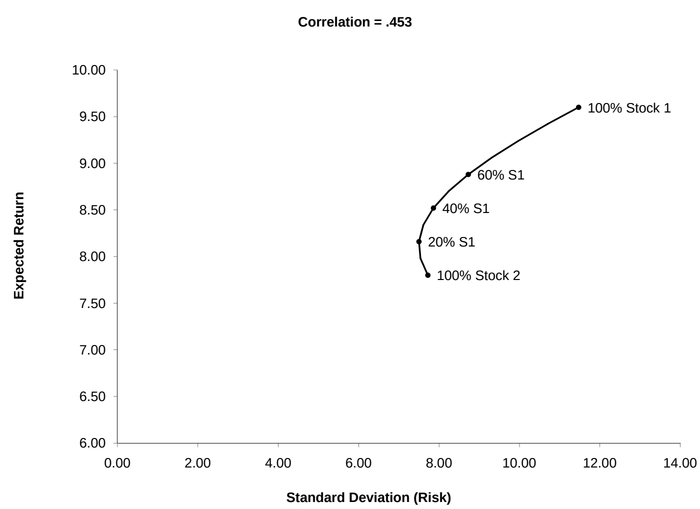


Figure 1 shows all of the possible combinations of expected return and standard deviation depending on what percentage of the portfolio consists of stock 1 or stock 2. The shape of it is fairly standard, but is dependent upon the correlation between the two stocks, which in this case was 0.453.

If we were to do the same calculations for the same two assets but assume that the correlation is either $\rho = +1$ or $\rho = -1$, the combinations would have a much different appearance. Figures 2 on the facing page and 3 on the next page show the similar calculation for $\rho = +1$ and $\rho = -1$.

FIGURE 2. Risk and Return — Two Assets with $\rho = 1$

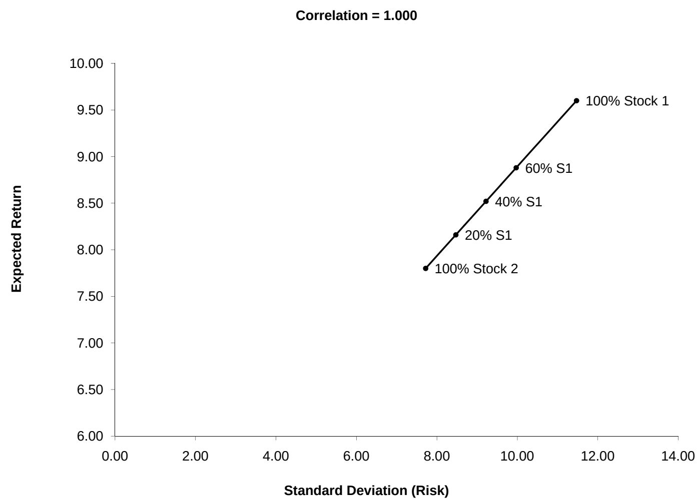
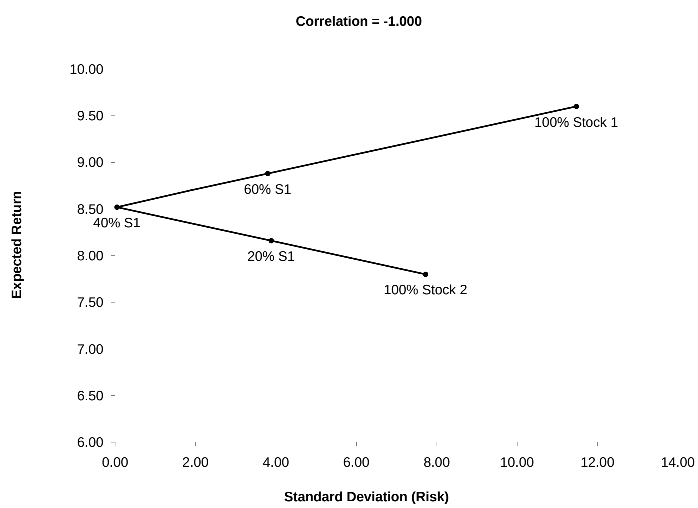


FIGURE 3. Risk and Return — Two Assets with $\rho = -1$



The differences between each of these cases arise solely from the degree of correlation between the two assets. When the correlation is $\rho = +1$, the portfolios all lie along a straight line; when the correlation is $\rho = -1$ they lie along a kinked line and one of them actually has a standard deviation of zero; and when the correlation is anywhere in between, there is a curved relationship. Higher standard deviation is associated with higher expected return, but the relationship is not linear.

Minimum Variance Portfolio

An important point about that is that if you look at the picture when $\rho = .453$, you will see that there is one portfolio in particular that has the lowest level of standard deviation compared to all of the others. In fact, it is even lower than the standard deviation of both of the assets in the portfolio.

This minimum standard deviation portfolio can be found using basic calculus by taking the first derivative of the portfolio standard deviation formula with respect to w_1 and setting this equal to zero. It turns out though that it is easier to minimize the variance rather than the standard deviation, so we will focus on that here. This will tell us how much to invest in each of w_1 and w_2 to achieve this minimum variance.

We know that the portfolio variance is given by the equation:

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho$$

If we want the minimum variance, then we simply need to take the derivative of this with respect to w_1, set the derivative equal to zero and solve for w_1.

We start by noticing that $w_2 = 1 - w_1$, which allows us to write the formula for the portfolio variance in terms of just w_1:

$$\sigma_p^2 = w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1(1 - w_1)\sigma_1\sigma_2\rho$$

and then the derivative is found as follows:

$$\begin{aligned}\frac{\partial(\sigma_p^2)}{\partial w_1} &= 2w_1\sigma_1^2 - 2(1 - w_1)\sigma_2^2 + 2\sigma_1\sigma_2\rho - 4w_1\sigma_1\sigma_2\rho \\ &= 2w_1(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho) - 2(\sigma_2^2 - \sigma_1\sigma_2\rho) \\ &= 0\end{aligned}$$

After some messy algebra, we can solve for w_1:

$$w_1 = \frac{\sigma_2^2 - \sigma_1\sigma_2\rho}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho}$$

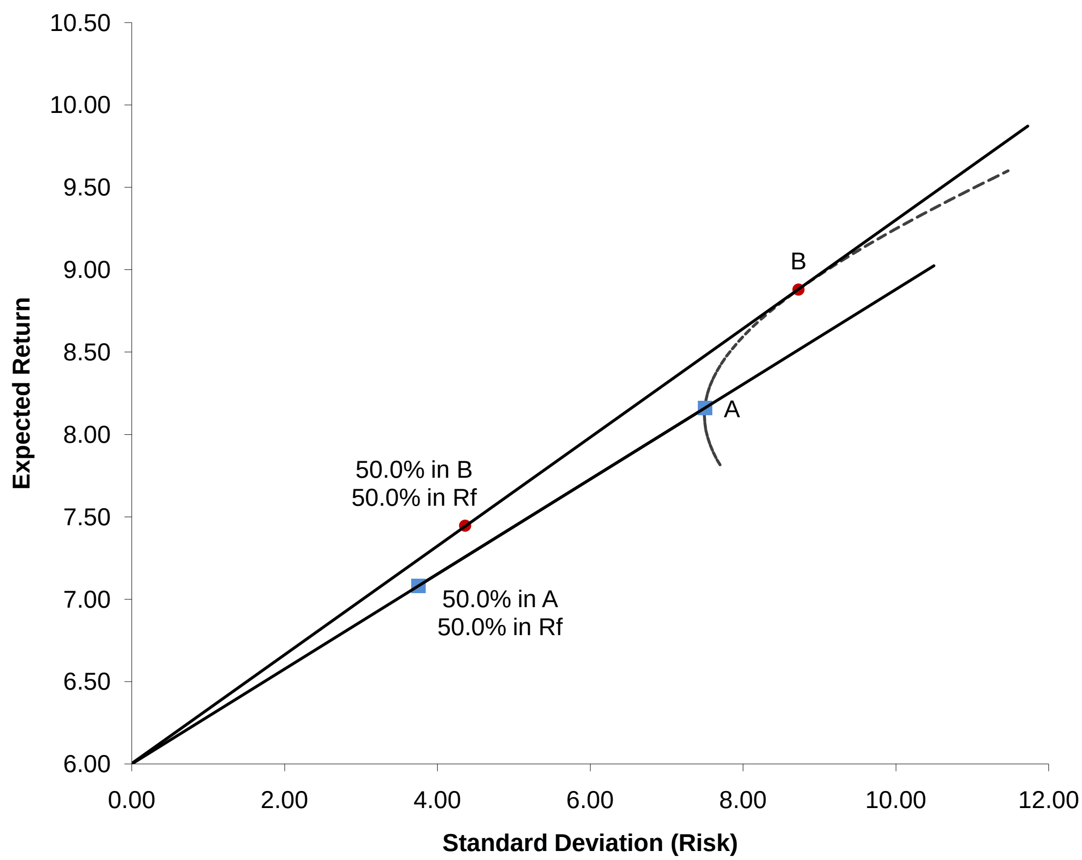
Asset Allocation With Risky Portfolio and Risk-Free Asset

Recall in Chapter 6 we looked at combinations of a risky portfolio and a risk-free asset. But now there are an infinite number of risky portfolios that we can combine with the risk-free assets, each one comprised of a different combination of the various risky assets.

Consider again the case with two risky assets with a correlation of $\rho = .453$. If we again draw the *portfolio opportunity* set depicting each of the possible combinations of the two risky assets, we see that we can select any of these points to be combined with the risk-free asset to achieve additional risk-return combinations.

Consider one of these points, A, along the portfolio opportunity set as well as combinations of this portfolio with the risk-free asset, $r_f = 6\%$, as shown in Figure 4. In addition to obtaining the risk and expected return depicted by A, combining this with the risk-free asset allows us obtain much lower risk or much higher risk portfolios than A simply by altering how much we allocate to A and to the risk-free asset.

FIGURE 4. Portfolio Opportunity Sets: Two Alternative Risky Assets



Is there another portfolio that would be even better? Yes, there is.

To find it, notice that all we want to do is keep finding lines that go through one of the risky portfolios and the risk-free point along the y-axis and that is higher than the other lines. Graphically, we simply want to find the tangent line to the portfolio opportunity set. The risky portfolio that produces this line is then the **optimal risky portfolio** and it is comprised of some optimal combination of the two risky assets.

Notice that the CAL with portfolio B is indeed the *optimal CAL*. It produces higher expected returns for any level of risk compared to the CAL obtained with portfolio A. From this, we can conclude that any investor attempting to maximize return and minimize risk will not find it optimal to combine the risk-free asset with portfolio A and instead would choose some combination of the risk-free asset and portfolio B.

Determining the Optimal Portfolio of Risky Assets

In mathematical terms, we want to find the combinations that produce the line with the highest slope (the highest Sharpe ratio). This is a simple maximization problem. We simply write the formula for the slope of that line and maximize it (recall from calculus that to maximize the function, take its first derivative, set that equal to zero and solve for the optimal weights).

The slope of the line, the Sharpe ratio, for any given portfolio, ρ, is simply:

$$\text{Sharpe Ratio: } S_p = \frac{E(r_p) - r_f}{\sigma_p}$$

The numerical problems below will walk you through this process.

Optimal Weight (Two Risky Assets Only)

The textbook gives the following formula for the weight on Risky Asset 1 that produces the optimal risky portfolio:

$$w_1 = \frac{[E(r_1) - r_f]\sigma_2^2 - [E(r_2) - r_f]\sigma_{12}}{[E(r_1) - r_f]\sigma_2^2 + [E(r_2) - r_f]\sigma_1^2 - [E(r_1) - r_f + E(r_2) - r_f]\sigma_{12}}$$

Note that this formula is not of general importance and only applies to the special case of two risky assets. It differs from the similar looking formula discussed earlier for the optimal weight for the *Minimum Variance Portfolio* because that formula was for the combinations of two risky assets that produce the minimum variance. This formula is for the combination of two risky assets that when combined with risk-free borrowing or lending produces the most efficient risk-return tradeoffs.

Memorization Tip: Notice that the formula for the weight in the optimal risky portfolio with borrowing and lending is very similar to the formula for the optimal weight in the minimum variance portfolio, except that the risk premiums $[E(r_i) - r_f]$ appear in front of the variance and covariance terms. Note also that the risk premiums for Asset 1 appear along with the variances for Asset 2, and vice versa. For the covariances term in the denominator, the risk premiums are added together. Otherwise, the terms are very similar:

Weight in Minimum Variance Portfolio:

$$w_1 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_2^2 + \sigma_1^2 - 2\sigma_{12}}$$

Weight in Optimal Risky Portfolio:

$$w_1 = \frac{[E(r_1) - r_f]\sigma_2^2 - [E(r_2) - r_f]\sigma_{12}}{[E(r_1) - r_f]\sigma_2^2 + [E(r_2) - r_f]\sigma_1^2 - [E(r_1) - r_f + E(r_2) - r_f]\sigma_{12}}$$

Derivation Using Simple Calculus

You should know the above formula (it has been asked on previous exams). It will be easier to make sense of this—and remember it—once you understand the derivation. The derivation itself will of course not be asked, so you don't have to worry about the tedious aspects of it. Just be vaguely familiar with the process because it is the general process that is important.

First, the book gives a simple formula for the Sharpe ratio (slope of the CAL) as:

$$S = \frac{E(r_p) - r_f}{\sigma_p}$$

This slope indicates the tradeoff between risk and return for any selected portfolio. To determine the value of w_1 which will maximize this slope (i.e. give the most excess return per unit of risk), we simply plug in our formulas for the expected return and standard deviation in terms of the unknown w_1, take the derivative with respect to w_1, set this to zero and solve for w_1.

To keep it simple, we will again substitute $w_2 = 1 - w_1$ so that we can take the derivative with respect to w_1 only.

$$E(r_p) = w_1 E(r_1) + (1 - w_1)E(r_2)$$

$$\begin{aligned}\sigma_p^2 &= w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1(1 - w_1)\sigma_1\sigma_2\rho \\ \sigma_p &= [w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1(1 - w_1)\sigma_1\sigma_2\rho]^{\frac{1}{2}}\end{aligned}$$

$$\begin{aligned}S &= \frac{E(r_p) - r_f}{\sigma_p} \\ &= \frac{w_1 E(r_1) + (1 - w_1)E(r_2) - r_f}{[w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1(1 - w_1)\sigma_1\sigma_2\rho]^{\frac{1}{2}}}\end{aligned}$$

The rest of the derivation is to simply take the derivative with respect to w_1, set it equal to zero and solve for w_1. After some messy algebra, the solution to the equation is simply:

$$w_1 = \frac{[E(r_1) - r_f]\sigma_2^2 - [E(r_2) - r_f]\sigma_{12}}{[E(r_1) - r_f]\sigma_2^2 + [E(r_2) - r_f]\sigma_1^2 - [E(r_1) - r_f + E(r_2) - r_f]\sigma_{12}}$$

This formula may look intimidating, but if you rewrite it in terms of the risk premiums with the substitution, $E(R_i) = E(r_i) - r_f$, then the formula is a little cleaner:

$$w_1 = \frac{E(R_1)\sigma_2^2 - E(R_2)\sigma_{12}}{E(R_1)\sigma_2^2 + E(R_2)\sigma_1^2 - [E(R_1) + E(R_2)]\sigma_{12}}$$

Finally, once we know the composition of the optimal risky portfolio and the equation for the optimal CAL, we can use our earlier results from maximizing utility to determine any given investor's preferred weighting between the risk-free asset and this optimal risky portfolio.

Markowitz Portfolio Selection Model

Now we will extend the above analysis to the more realistic case when there are more than two risky assets to choose from. As we will see, the differences are quite minimal.

Risk and Return for the Risky Asset Portfolio

The first step is to determine all of the combinations of risk and return for the various combinations of the risky asset. This is the exact parallel to depicting the portfolio opportunity set for two assets.

Consider first the case with three assets. Here, the formulas for expected return and portfolio variance are similar to the two portfolio case:

$$E(r_p) = w_1E(r_1) + w_2E(r_2) + w_3E(r_3)$$

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_3^2\sigma_3^2 + 2w_1w_2\sigma_{12} + 2w_1w_3\sigma_{13} + 2w_2w_3\sigma_{23}$$

The variance formula contains a term like $w_i^2\sigma_i^2$ for each asset and a term like $2w_iw_j\sigma_{ij}$ for each pair of assets (1 and 2, 1 and 3, 2 and 3).

If we then extend this to N assets, we have the following general formulas:

$$E(r_p) = \sum_{i=1}^{n} w_iE(r_i)$$

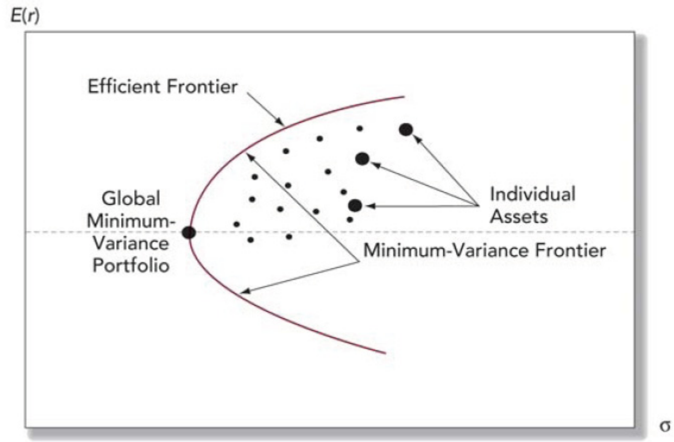
$$\sigma_p^2 = \sum_{i=1}^{n} w_i^2\sigma_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} w_iw_j\sigma_{ij}$$

Notice that the double summation symbol picks up every pair so long as $i \neq j$.

Each of these combinations of risk and return are then graphed in the same way we did it for the two asset case. As shown in Figure 5 on the next page there will be an infinite number of points, but there will be an outer frontier that will look much like what we had before, with each point along the frontier representing the minimum risk portfolio for any level of expected return.

Notice that again there is a single portfolio combination that results in the Global Minimum Variance. Note also that all points along this frontier are possible, but only those above this

FIGURE 5. Minimum Variance Frontier



global minimum variance portfolio are of interest because the points below are *inefficient*. They each have lower expected returns for any level of risk than some other point above the global minimum variance portfolio.

For this reason, the points above the global minimum are referred to as *efficient portfolios* or collectively as the *efficient frontier*. These points represent the portfolios that have the highest return for any level of risk and the lowest risk for any level of return. These efficient portfolios are the only portfolios that investors would choose from out of all of the possible portfolios.

Optimal Risky Portfolio with Many Risky Assets

Once the efficient portfolios are identified, the next step is to determine the single portfolio that when combined with the risk-free asset produces the best possible risk and return combinations. This optimal portfolio is found in the same fashion as before — we write the formula for the slope of the Capital Allocation Line and maximize this function.

Here though, there will be multiple assets to deal with and therefore we have two mathematical complications:

1. We need to take the (partial) derivative with respect to each asset's weight, set it equal to zero and then solve each of these equations simultaneously.

2. The optimization problem is further complicated by the constraint that the sum of all the weights must add to 1.0. If we do not allow the weights to be less than zero (i.e. we do not allow short sales) then we may have additional constraints as well.

The text describes how this would be done using Excel's Solver rather than using linear algebra and calculus. It also shows the resulting formula for the special case where there are only two risky assets.

Capital Allocation Line (CAL)

Once the optimal risky portfolio is found, the formula for the CAL can be written, just as was done before with two assets.

Allocation

Once the CAL is determined, each investor's utility function is used to determine their optimal allocation between this optimal risky portfolio and the risk-free asset.

Example

It may be useful to see a bit of a demonstration of this.

Assume that you wanted to find the optimal risky portfolio and the CAL from a universe of seven risky assets. With seven assets there will be a lot of possible combinations of these that given different risk and return combinations but one of these will produce a CAL that produces risk-return combinations that are better than any other point.

Let's see if this really works with specific assumptions, including $r_f = 3\%$ and the assumptions shown in Table 1 and Table 2.

TABLE 1. Risk and Return Characteristics

Risky Asset	E(r)	Std Dev
1	9.0%	13.6%
2	16.7%	38.2%
3	9.4%	29.0%
4	6.2%	19.4%
5	11.4%	26.1%
6	7.0%	18.2%
7	7.3%	19.9%

TABLE 2. Correlation Matrix

Risky Asset	1	2	3	4	5	6	7
1	1.00	0.72	0.58	0.43	0.66	0.35	0.46
2	0.72	1.00	0.46	0.10	0.43	0.12	0.29
3	0.58	0.46	1.00	0.38	0.45	-0.02	0.13
4	0.43	0.10	0.38	1.00	0.62	-0.01	0.01
5	0.66	0.43	0.45	0.62	1.00	0.14	0.16
6	0.35	0.12	-0.02	-0.01	0.14	1.00	0.74
7	0.46	0.29	0.13	0.01	0.16	0.74	1.00

In Panel (A) of Figure 6 on the next page, I show the seven individual assets (denoted as black squares), the optimal portfolio (denoted with a white circle outlined in red) calculated using the procedure described above and the capital allocation line when this optimal portfolio is combined with the risk-free asset ($r_f = 3\%$).

I then add twenty randomly generated portfolios with different weights on each of the seven risky portfolios in Panel (B). While some of these may have more appealing risk and return relationships than some of the assets, none are above the capital allocation line.

Panels (C) and (D) are similar but contain 500 and 2,000 random portfolios, confirming that the optimal portfolio is indeed optimal and tangent to the efficient frontier depicted by these 2,000 portfolios.

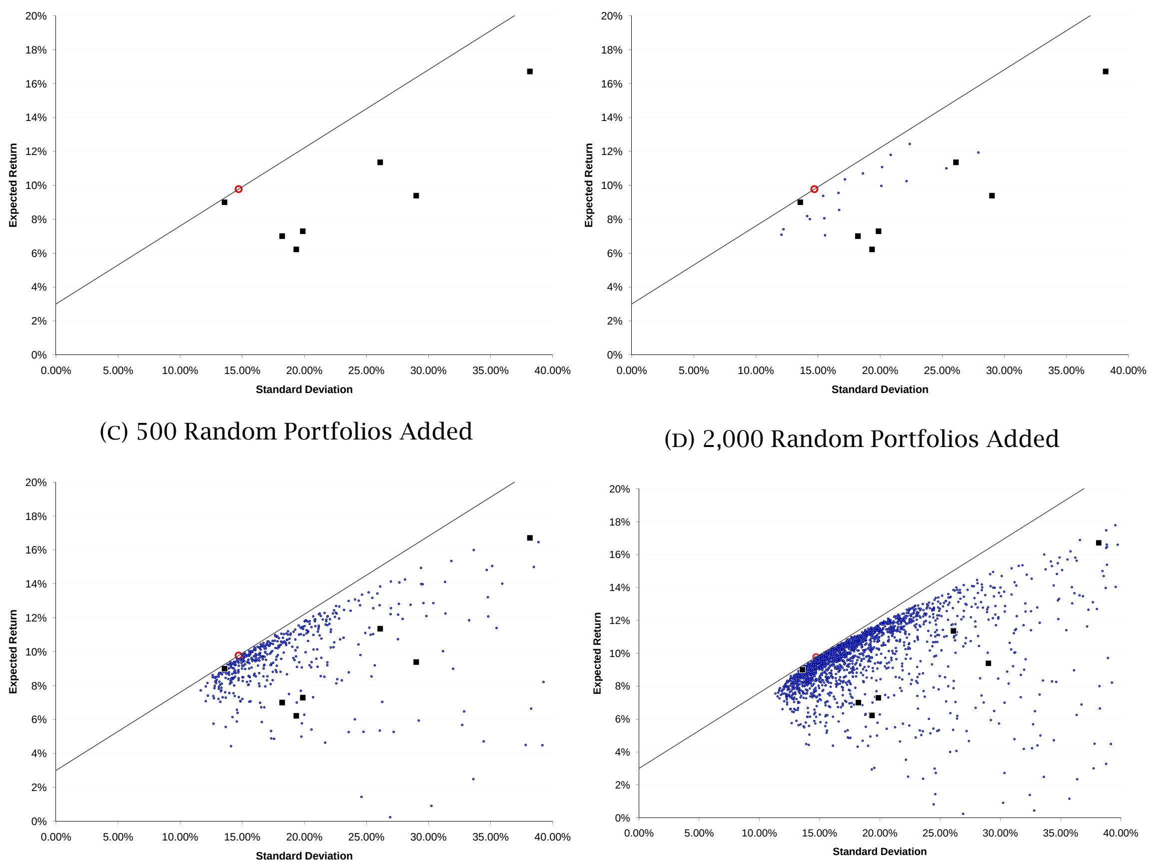
FIGURE 6. Optimal Risky Portfolio with Multiple Risky Assets

(A) Risky Assets and Optimal Portfolio

(B) 20 Random Portfolios Added

(C) 500 Random Portfolios Added

(D) 2,000 Random Portfolios Added



Separation Property

The description above makes it clear that portfolio selection can be separated into two very distinct tasks.

1. The first is to select an optimal combination of risky assets. This is a purely mathematical exercise, given the assumptions of expected return, standard deviation and correlations for each of the risky assets, and is not affected by any individual's risk preferences.

2. The second task is to allocate funds between the risk-free asset and the optimal risky portfolio. This is entirely based upon individual risk preferences.

The Power of Diversification

Earlier, we showed that when there are multiple risky assets in a portfolio,

$$\sigma_p^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} w_i w_j \sigma_{ij}$$

Considering the more general case with N assets allows us to make an important observation about the variance of returns for a portfolio. Assume that we have N assets in the portfolio and that we have equal amounts invested in each, such that each $w_i = 1/N$. If we plug those amounts in for the weights and rearrange things a bit, we get the following:

$$\sigma_p^2 = \frac{1}{N} \underbrace{\sum_{i=1}^{n} \frac{\sigma_i^2}{N}}_{\text{Avg. Variance}} + \frac{N-1}{N} \underbrace{\sum_{i=1}^{n} \sum_{j \neq i} \frac{\sigma_{ij}}{N(N-1)}}_{\text{Avg. Covariance}}$$

Notice that the first summation term represents the average of each of the variances for the N assets. Similarly, the second double summation term represents the average covariance of each pair of assets. So we can rewrite the formula for the portfolio variance when there are N assets each with a weight equal to $w_i = 1/N$ as:

$$\sigma_p^2 = \frac{1}{N} * \text{Avg. Variance} + \frac{N-1}{N} * \text{Avg. Covariance}$$

This formula is important. It shows that as N gets large (i.e. as we increase the number of assets in our portfolio and continue to invest an equal amount in each asset) the first term becomes very small since the ratio $1/N$ approaches zero and the second term approaches the average covariance since the ratio $(N-1)/N$ approaches 1. This means that the variance of the portfolio as N gets large depends primarily on the covariances of the assets in the portfolio and not their respective variances.

Risk Pooling vs. Risk Sharing

The previous discussion of diversification suggests that in the special case where all of the assets have the same standard deviation and correlations, as N gets larger the portfolio standard deviation declines. Many people use that rationale to explain the existence of insurance companies—they say that as the insurer takes on more and more risks, diversification reduces their total risk.

That seemingly logical statement is actually incorrect. As insurers write more policies they take on more absolute risk in dollar terms. What is true though is that their risk as a percentage of the value of the policies they write declines. This is because the total risk increases in proportion to the square root of N, where N is the number of equal sized risks, while the value increases in proportion to N. Another way to say this is that the risk-reward tradeoff improves (the Sharpe ratio increases) with diversification but the standard deviation, and thus the total risk, increases.

Notice the important difference between increasing the number of assets in your portfolio and adding more risks in the insurance company context. In the first case, as we increase N, we have also assumed that the percentage held in each asset declines to $1/N$. In other words we held our total investment amount constant. Therefore, the simple pooling of risks is not sufficient to explain the existence of insurance companies. In the insurance context, adding more risks increases the total size of the portfolio. The only way for diversification to decrease risk is if your portfolio reflects smaller percentages of a larger number of independent risks. That is, there must be risk sharing.

Risk in the Long Run

The concepts of risk pooling and risk sharing discussed above apply equally well to an often misunderstood concept commonly referred to as the risk in the long run. For instance, it is often said that while stocks are risky, they are less risky in the long run because *time diversification* causes the average long-term rate of return to be more stable. While this is true mathematically, it ignores the fact that the dollar value of the risky bets is increasing over time. When the risk is measured in terms of dollars, such as the dollar amount of a shortfall at some future point in time, it becomes clear that risk increases over time—just not in proportion to the length of time.

Practice Questions

Question 1. Suppose you have two stocks, S_1 and S_2. The expected returns are 9.6% and 7.8%. The standard deviations are 11.47% and 7.72%. The correlation between the two stocks is .453. What are the expected return and standard deviation of a portfolio invested 40.2% in S_1 and 59.8% in S_2?

Note: Feel free to round to 40% and 60% for the weights, but my answers were not rounded.

Solution.

$$\begin{aligned}
E(r_p) &= w_1 E(r_1) + w_2 E(r_2) \\
&= .402(9.6) + .598(7.8) \\
&= 8.52 \\
\sigma_p^2 &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho \\
&= .402^2(11.47)^2 + .598^2(7.72)^2 + 2(.402)(.598)(11.47)(7.72)(.453) \\
&= 61.78 \\
\text{Std. Dev.} &= (61.78)^{1/2} \\
&= 7.86
\end{aligned}$$

To ease the notation the book drops the % signs in the expected return and variance calculations. The units for the variance is %-squared so when I take the square root in the standard deviation calculation the units are back to %. This is generally okay, but for questions related to utility functions it is important to use the decimal values (.1147 and .0772, for instance).

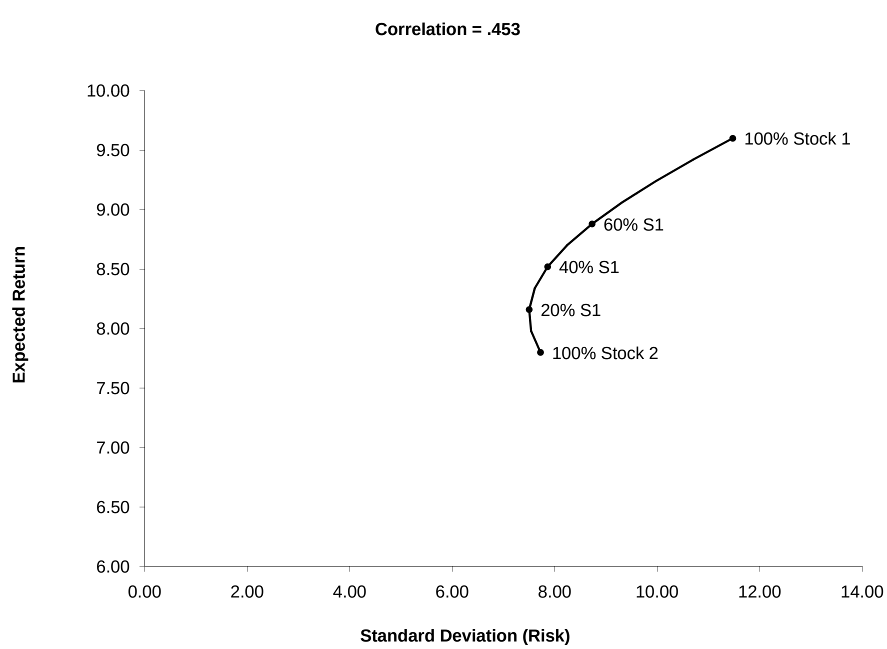
Question 2. Suppose you have two stocks, S_1 and S_2. The expected returns are 9.6% and 7.8%. The standard deviations are 11.47% and 7.72%. The correlation between the two stocks is .453. The following table shows the expected return and standard deviations for portfolios using three different sets of weights for the two stocks.

w_1	w_2	$E(r_p)$	σ_p
0.20	0.80	8.16	7.50
0.40	0.60	8.52	7.86
0.60	0.40	8.88	8.73

Graph these risk and return combinations along with the values you would obtain with 100% in Stock 1 or 100% in Stock 2.

Solution. The graph is the same one shown in the notes:

FIGURE 8. Risk and Return



Question 3. Suppose you have two stocks, S_1 and S_2. The expected returns are 9.6% and 7.8%. The standard deviations are 11.47% and 7.72% and the correlation is -1.0. Calculate the expected return and standard deviation of a portfolio invested 40.2% in S_1 and 59.8% in S_2.

Solution. This is just a matter of redoing the same math with the different assumptions regarding the correlation coefficient. Because the returns are perfectly negatively correlated, with these particular weights the standard deviation winds up being zero.

$$\begin{aligned}
E(r_p) &= w_1 E(r_1) + w_2 E(r_2) \\
&= .402(9.6) + .598(7.8) \\
&= 8.52 \\
\sigma_p^2 &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho \\
&= .402^2(11.47)^2 + .598^2(7.72)^2 + 2(.402)(.598)(11.47)(7.72)(-1) \\
&= 0 \\
\text{Std. Dev.} &= (0)^{1/2} \\
&= 0
\end{aligned}$$

Question 4. Suppose you have two stocks, S_1 and S_2. The expected returns are 9.6% and 7.8%. The standard deviations are 11.47% and 7.72%. The correlation between the two stocks is .453. The following table shows the expected return and standard deviations for portfolios using two different sets of weights for the two stocks.

	w_1	w_2	$E(r_p)$	σ_p
Portfolio A	0.20	0.80	8.16	7.50
Portfolio B	0.60	0.40	8.88	8.73

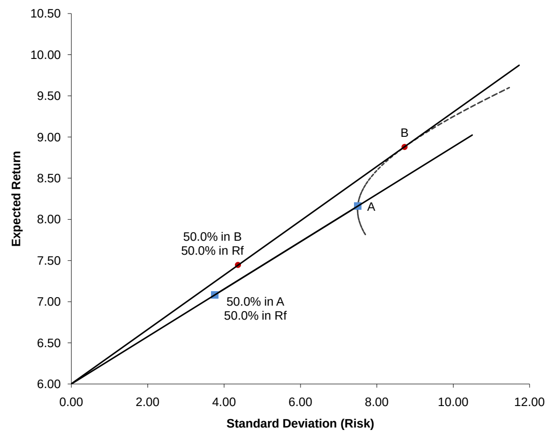
Separately for each of these portfolios, plot the points representing three different combinations of these portfolios with the risk-free asset with an expected return of 6% (and a standard deviation of zero). Use w_{RF} equal to 100%, 50% and 0%.

Solution. This question seems rather convoluted, but it highlights the most important point of the entire chapter. I won't show the steps, but the table below summarizes the results:

Risky Portfolio A			Risky Portfolio B		
w_{RF}	$E(r_c)$	σ_c	w_{RF}	$E(r_c)$	σ_c
100%	6.00	0.00	100%	6.00	0.00
50%	7.08	3.75	50%	7.44	4.36
0%	8.16	7.50	0%	8.88	8.73

Now, plot each of these points on the same graph and draw a line depicting the combinations of each risky portfolio with the risk-free asset. For clarity, I will show this with the y axis starting at 6%:

FIGURE 9. Possible Complete Portfolios

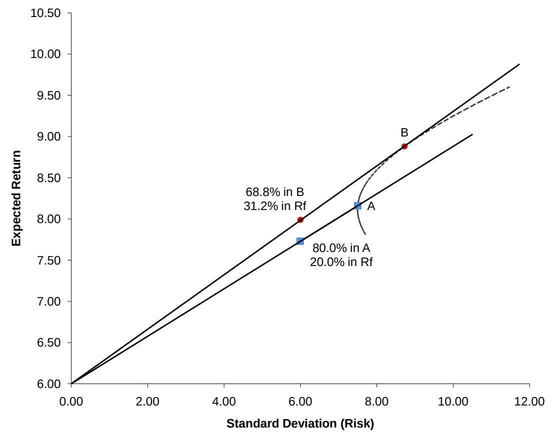


Question 5. Refer to the graph in the previous Question. If you wanted to invest in a portfolio that had a standard deviation of 6.0%, which is the best way to accomplish this?

Solution. Notice that to achieve a portfolio with 6% standard deviation, we could combine the risk-free asset with either portfolio. However, if we combined it with the portfolio that has 60% in Stock 1 and 40% in Stock 2, the higher line in the picture, then we could have a 6% standard deviation and an expected return of approximately 8%. This is better than combining it with the other portfolio because in that case you would get a 6% standard deviation but the expected returns would be just over about 7.7%.

Also note that the 8.25% expected return that could be achieved here could also be achieved without the risk-free asset at all — it was one of the points on the original graph. In that case, the standard deviation would be in the range of 7.5% — much higher than the 6% that we would have if we used the risk-free asset.

FIGURE 10. Possible Complete Portfolios with Std. Dev = 6%



Notice too that there is no portfolio on the efficient frontier that either on its own or combined with the risk-free asset that can achieve any target risk or return more efficiently. Portfolio B then is said to be the *optimal risky portfolio* when borrowing and lending is allowed.

Question 6. Suppose an investor wanted to select his optimal complete portfolio consisting of the two stocks used in the previous questions, with expected returns of 9.6% and 7.8%, respectively, standard deviations of 11.47% and 7.72%, respectively, and correlation coefficient of .453. Further assume that the investor's utility function is of the form $U = E(r) - .5A\sigma^2$ and $A = 3$.

What would be the weights for each stock if these are the only two assets to choose from and risk-free borrowing or lending are not allowed?

Solution. As we did in the previous chapter, here we simply assume that the investor will want to maximize his utility. We know that the expected return of his portfolio will be:

$$E(r) = .096w_1 + .078(1 - w_1)$$

The variance of his portfolio will be:

$$\sigma^2 = w_1^2(.1147)^2 + (1 - w_1)^2(.0772)^2 + 2w_1(1 - w_1)(.1147)(.0772)(.453)$$

Now, we just plug these formulas into the formula for utility, take the derivative with respect to w_1 and set it equal to zero.

When this is done (it's a mess!), you can solve for $w_1 = .717$. The following table demonstrates that this is in fact the weight that maximizes the utility.

TABLE 3. Utility for Different Weights on Stock 1

w_1	Utility
0.000	0.069060
0.100	0.071278
0.200	0.073164
0.300	0.074716
0.400	0.075936
0.500	0.076823
0.600	0.077377
0.700	0.077599
0.717	**0.077603**
0.800	0.077487
0.900	0.077043
1.000	0.076266
1.100	0.075156
1.200	0.073713
1.300	0.071938
1.400	0.069830
1.500	0.067389

Question 7. You have two stocks, Stock 1 and Stock 2, and want to find the weights w_1 and w_2 of the *optimal risky portfolio*—the portfolio of risky assets 1 and 2 that, when combined with the risk-free asset, produce complete portfolios with the highest possible returns for any level of risk.

For Stock 1, the expected return is 9.6% and the standard deviation is 11.47%. For Stock 2 the expected return is 7.8% and the standard deviation is 7.72%. The correlation of returns between Stock 1 and Stock 2 is .453. Assume that the risk-free rate is 6%.

Solution. Recall what we did in the notes. We first selected a particular portfolio of risky assets and determined its expected return and standard deviation. Then we combined this with the risk-free asset and drew the line connecting the risk-free asset and the risky portfolio. Now, we just want to pick the one portfolio that causes the slope of this line to be the greatest.

The formula given in the textbook for the case where there are only two risky assets was:

$$w_1 = \frac{[E(r_1) - r_f]\sigma_2^2 - [E(r_2) - r_f]\sigma_{12}}{[E(r_1) - r_f]\sigma_2^2 + [E(r_2) - r_f]\sigma_1^2 - [E(r_1) - r_f + E(r_2) - r_f]\sigma_{12}}$$

Plugging in the specific values for this problem,

$$\begin{aligned} w_1 &= \frac{(3.6)(7.72^2) - (1.8)(11.47)(7.72)(.453)}{(3.6)(7.72^2) + (1.8)(11.47^2) - (3.6 + 1.8)(11.47)(7.72)(.453)} \\ &= .6064 \\ &= 60.64\% \\ w_2 &= 39.36\% \end{aligned}$$

These were (approximately) the weights used for Portfolio B in the graph we plotted in an earlier question. When combining the risk-free asset with this portfolio, the various risk and return combinations lie along the straight line that is *tangent* to the efficient frontier.

Also note that the formula for the optimal weight shown in the textbook and used here applies only to the special case with only two risky assets. For the more general result with multiple risky assets you would actually have to determine the weights that produce the maximum Sharpe ratio using calculus.

Question 8. In the previous question, I mentioned drawing the line that is tangent to the efficient frontier. What is the algebraic equation for this line?

Solution. The equation for a line in *slope-intercept* form is $Y = mX + b$. Here, $Y = E(r)$, $X = \sigma$, m = slope = $[r_p - r_f]/\sigma_p$ and $b = r_f$. Therefore, the equation for the line is:

$$E(r) = r_f + \frac{r_p - r_f}{\sigma_p}\sigma$$

In this case, we know that the risky portfolio consists of 61% invested in Stock 1 and 39% invested in Stock 2, and so we know that this portfolio has an expected return of 8.898% and a standard deviation of 8.781%.

Therefore, the CAL can be written as:

$$E(r) = .06 + \frac{.08898 - .06}{.08781}\sigma = .06 + .33\sigma$$

Question 9. Referring to the situation described in the previous three questions, suppose the investor can now combine risk-free borrowing or lending with his optimal risky asset portfolio consisting of Stocks 1 and 2. What will be his optimal portfolio composition?

Solution. First, we'll use the CAL from the previous question and the investor's utility curve to determine his optimal allocation, y^*, to the risky portfolio.

Here, the expected return is as specified by the CAL ($E(r) = .06 + .33\sigma$) and the standard deviation of his portfolio is $\sigma = y\sigma_p = y(.08781)$.

Combining these, we get $U = .06 + .33y(.08781) - .5(3)(y^2)(.08781)^2$.

Taking the derivative and setting it equal to zero we get:

$$0 = .33(.08781) - 2(.5)(3)(y)(.08781^2)$$

Solving for y^* gives us:

$$y^* = \frac{.33(.08781)}{2(.5)(3)(.08781^2)} = 1.253$$

This means that this particular investor, with risk aversion $A = 3$, will invest 125.3% of his assets in the risky portfolio, which itself consists of 60.64% in Stock 1 and 39.36% in Stock 2. The final composition of the portfolio will be 76.4% in Stock 1, 48.6% in Stock 2 and borrowing 25.3% at the risk-free rate.

Question 10. Suppose you have a portfolio consisting of two stocks, S_1 and S_2. Assume that the expected returns are 9.6% and 7.8%, respectively. The standard deviations are 11.47% and 7.72% and the correlation is .453. Determine the proportion of your portfolio you would invest in S_1 and S_2 such that you had the minimum possible portfolio variance.

Solution. First I will show the derivation in symbols and then in the final step I will plug in the values given.

We know from the work we've done already that the portfolio variance is given by the equation:

$$\sigma_p^2 = w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1(1 - w_1)\sigma_1\sigma_2\rho$$

and then the derivative is found as follows:

$$\begin{aligned}
\frac{\partial(\sigma_p^2)}{\partial w_1} &= 2w_1\sigma_1^2 - 2(1 - w_1)\sigma_2^2 + 2\sigma_1\sigma_2\rho - 4w_1\sigma_1\sigma_2\rho \\
&= 2w_1(\sigma_1^2 + \sigma_2^2 - \sigma_1\sigma_2\rho) - 2(\sigma_2^2 - \sigma_1\sigma_2\rho) \\
&= 0 \\
w_1 &= \frac{\sigma_2^2 - \sigma_1\sigma_2\rho}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho} \\
&= \frac{7.72^2 - (11.47)(7.72)(.453)}{11.47^2 + 7.72^2 - 2(11.47)(7.72)(.453)} \\
&= .1757 \\
w_2 &= .8243
\end{aligned}$$

Question 11. Describe how would you find the minimum variance portfolio as we did in the previous question if you had three or more stocks in the portfolio?

Solution. In this case, you would have one equation but with two or more unknowns (before because $w_2 = 1 - w_1$ we only had one unknown). To get the minimum, you would have to take the derivative with respect to each variable and set them each equal to zero simultaneously. This gives you a system of linear equations to solve simultaneously.

Question 12. Suppose the universe of risky assets consists of a large number of stocks, each of which has an expected return of 12%, a standard deviation of 50% and a common correlation coefficient of .4. The risk-free rate is 4%. How much do the portfolio standard deviation and the portfolio Sharpe ratio fall when moving from a portfolio entirely invested in just one stock to a portfolio with equal amounts invested in twenty stocks?

Solution. With just one stock, the standard deviation is 50% and the Sharpe ratio is:

$$S = \frac{E(r_p) - r_f}{\sigma_p} = \frac{.12 - .04}{.50} = .16$$

With twenty stocks, note that the average covariance is $\rho * \sigma^2 = .4(.5^2) = .1$. Then, the portfolio variance becomes:

$$\begin{aligned}\sigma_p^2 &= \frac{1}{N} * \text{Avg. Variance} + \frac{N-1}{N} * \text{Avg. Covariance}\\ &= \frac{1}{20} * (.5^2) + \frac{19}{20} * (.1)\\ &= .1075\end{aligned}$$

Then, the standard deviation is $\sqrt{.1075} = 0.328$. Notice that this is about 65% of the portfolio standard deviation when there was just one stock.

For the Sharpe ratio, notice that the expected risk premium is the same because the expected returns are the same and when moving from one stock to twenty stocks we held the amount invested constant. Therefore, the Sharpe ratio is:

$$S = \frac{E(r_p) - r_f}{\sigma_p} = \frac{.12 - .04}{.328} = .24$$

Through diversification we were able to reduce the portfolio risk and increase the Sharpe ratio.

Question 13. Using the same information as above, what is the systematic risk in this security universe?

Solution. When N gets very large, the portfolio variance will approach the average covariance, or $\rho * \sigma^2 = .4(.5^2) = .1$, and therefore the portfolio standard deviation will approach $\sqrt{.1} = .316$.

Question 14. An investor's complete portfolio, denoted P, consists of 60% invested in a single risky asset, A, and the remaining 40% invested in the risk-free asset, with $r_f = 4\%$. The risky asset has an expected return of 10% and a standard deviation of 40%. He now considers adding another independent risky asset, B, with the same expected return and standard deviation to this portfolio and plans to invest the same amount in this new asset as he already has invested in A. What happens to the standard deviation and the Sharpe ratio of his new complete portfolio, Z, after adding B?

Solution. Recall from the discussion of complete portfolios that if y is the percent invested in the risky asset than the portfolio risk premium is simply $R_P = y[E(r_A) - r_f] = .6(6\%) = 3.6\%$ and the portfolio standard deviation is $\sigma_P = y\sigma_A = .6(.4) = 24\%$. This makes the original portfolio Sharpe ratio equal to $S_P = 3.6\%/.24 = 15\%$.

By adding the same amount of B as originally invested in A, we now have $y = 60\%$ invested in A, another $y = 60\%$ invested in B and $1 - 2y = -20\%$ invested in the risk free asset. We have to borrow at the risk free rate to invest in B.

Now, the risk premium for the complete portfolio is $R_Z = yR_A + yR_B + (1 - 2y)(0) = 7.2\%$.

Similarly, the portfolio standard deviation is:

$$\begin{aligned}
\sigma_Z &= \sqrt{y^2\sigma_A^2 + y^2\sigma_B^2 + (1-2y)^2(0)} \\
&= \sqrt{.6^2(.4^2) + .6^2(.4^2)} \\
&= 34\%
\end{aligned}$$

Finally, the Sharpe ratio is $S_Z = 7.2\%/34\% = .212$.

Notice that the effect of diversification does not reduce the standard deviation of the complete portfolio, it actually rises from 24% to 34%. However, diversification does improve the Sharpe ratio. This demonstrates the effect of *risk pooling*.

Question 15. Using the same assets as in the previous question, show the effect of diversification if the investor were to hold the allocation to risky assets constant at 60% of the portfolio value.

Solution. Here, the difference is that the investor will continue to have 40% invested in the risk free asset and will now split his investment so that 30% is invested in A and 30% is invested in B. This changes the calculations as follows:

$$\begin{aligned}
R_Z &= .3(.06) + .3(.06) + .4(0) = 3.6\% \\
\sigma_Z &= \sqrt{.3^2(.4^2) + .3^2(.4^2)} = 17\% \\
S_Z &= 3.6\%/17\% = .212
\end{aligned}$$

In this case, because the dollar amount invested in the risky assets remained constant, the standard deviation decreased and the Sharpe ratio also increased. This demonstrates the effect of *risk sharing*.

Question 16. Suppose you were going to enter into a gamble such that there was a 60% chance of winning \$2 and a 40% chance of losing \$1. What is the mean and standard deviation of this gamble if you play once? What is the ratio of the standard deviation to the mean? What if you play ten times?

Solution. For a single play, the expected payoffs are $.6(2) + .4(-1) = .8$. The variance is:

$$\sum p_i(x-\mu)^2 = .6(2-.8)^2 + .4(-1-.8)^2 = 2.16$$

and the standard deviation is then 1.47. The ratio of standard deviation over the mean is $1.47/.8 = 1.84$.

For multiple independent plays, the mean is $N *$ Mean and the Variance is $N *$ Variance. So here, the mean is $10 * .8 = 8$ and the variance is $10 * 2.16 = 21.6$. The standard deviation is then 4.65. The ratio of standard deviation over the mean is .58.

Question 17. Suppose you had a portfolio of three stocks A, B and C. The expected returns and standard deviations of returns for each of the three stocks are as follows:

TABLE 4. Risk and Return for Stocks A, B and C

Stock	$E(r)$	Standard Deviation
A	14%	12%
B	16%	20%
C	13%	15%

Assume the correlation of returns for each pair of stocks is as follows $\rho_{AB} = .45$, $\rho_{AC} = .2$ and $\rho_{BC} = .6$. What are the expected returns and standard deviations for a portfolio of the three with 30% in Stock A, 50% in Stock B and 20% in Stock C?

Solution. The formulas are just an extension of what we did before for two stocks:

$$\begin{aligned} E(r_p) &= w_1 E(r_1) + w_2 E(r_2) + w_3 E(r_3) \\ &= .3(14) + .5(16) + .2(13) \\ &= 14.80 \end{aligned}$$

$$\begin{aligned} \sigma_p^2 &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_3^2\sigma_3^2 \\ &\quad + 2w_1w_2\sigma_{12} + 2w_1w_3\sigma_{13} + 2w_2w_3\sigma_{23} \\ &= .3^2(12^2) + .5^2(20^2) + .2^2(15^2) \\ &\quad + 2(.3)(.5)(12)(20)(.45) + 2(.3)(.2)(12)(15)(.2) \\ &\quad + 2(.5)(.2)(20)(15)(.6) \\ &= 194.68 \end{aligned}$$

$$\text{Std Dev} = \sqrt{194.68} = 13.95$$

The following series of questions relate to the material on Risk Pooling and Risk Sharing.

Question 18. Assume that an insurer is considering writing a policy with \$6,000 in premium and potential claim payments (a loss distribution) that are normally distributed with a mean of \$5,000 and a standard deviation of \$5,000. What is the dollar loss for the insurer from this policy at the 97.5th percentile?

Note: Recall that the 97.5 the percentile *loss* is equal to Mean Profit − 1.96 ∗ Std Deviation.

Solution. First, note that because of the \$6,000 to be paid in premiums, the dollar profit to the insurer is $6,000 - L$, where L is the amount of the claim. The expected value of this is $6,000 - 5,000 = 1,000$. Because the variance of the premium is zero, the variance of the profit is $\text{Var}(6,000 - L) = \text{Var}(L) = 5,000^2 = 25,000,000$. The standard deviation of the dollar profit is then \$5,000.

Using this, the 97.5th percentile loss is equal to $1,000 - 1.96(5,000) = -8,800$.

Question 19. Now consider writing 10 policies just like the one in the previous question and assume that the policies are independent. What is the insurer's expected profit and standard deviation? What is the 97.5th percentile loss?

Solution. Here, we simply note that because the different policies are independent, you get the expected value and variances by adding the individual means and variances. So expected profit is $10(1,000) = 10,000$. Variance = $10(25,000,000) = 250,000,000$. The standard deviation is then $5,000(10^{.5}) = 15,811$. The 97.5th percentile loss is -20,990.

Question 20. Assume that we define risk for the insurer in the previous question as the amount of money (capital) they would need to ensure that their losses did not exceed their capital more than 2.5% of the time. In this case, they would set their capital requirement equal to the 97.5th percentile loss. Does writing 10 policies rather than 1 policy reduce their risk as defined in this way? Assume each of the policies is independent.

Solution. Clearly, risk is not reduced by writing 10 policies, because in the first case the firm needed capital of $\$8,800$ and in the second it needed capital of $\$20,990$.

Question 21. Referring to the previous question, show that risk as defined in that question would in fact be reduced if instead they wrote 10 policies, each of which was only 1/10 the size of the original. Assume each of the policies is independent.

Solution. If they write 10 policies and each policy were 1/10 the size, the expected profit for each policy would be $(1/10)(1,000) = 100$.

Similarly, the variance would be $(1/10)^2(25,000,000) = 250,000$ and the standard deviation would be $\sqrt{250,000} = 500$.

Then if we had 10 of these, the expected profit would be $10(100) = 1,000$ and the standard deviation would be $\sqrt{10}(500) = 1,581$.

Finally the 97.5th percentile loss would be $-2,099$.

Notice that in this case, the expected value is the same as if we just wrote a single large policy, but now, the risk as defined this way is reduced considerably from $8,800$ to $2,099$. This demonstrates the point the book makes that risk needs to be measured in dollars, not rates of return, in order to evaluate whether diversification is occurring.

BKM Chapter 8: Index Models

One of the key weaknesses of the approach to portfolio theory we have presented so far is that the data needed to implement it is excessive. We need to know, for each possible risky asset, its expected return and its standard deviation, as well as the correlation coefficients for every pair of assets. That's an immense amount of data to estimate, so this chapter presents a method to simplify the estimation of all of these parameters.

Single Factor Model

The single factor model assumes that the return for a security is the sum of both expected and unexpected components. Further, we assume that the return deviates from its expected return for two reasons — unanticipated macroeconomic (economy-wide) events that affect all assets and unanticipated firm-specific events. If all of the macro events can be summarized using a single factor, m, and security i's sensitivity to that factor can be measured as β_i, then the return for security i can be written as:

$$r_i = E(r_i) + \beta_i m + e_i$$

Here, we are depicting the *actual* realized return for a specific security and suggesting that it reflects the security's expected return, a return driven by a macroeconomic factor common to all securities and a return due to some firm-specific sources of variability.

Variance of Returns

Using this equation for the single-factor model we can also depict the variance of returns for a security by taking the variance of the above expression:

$$\sigma_i^2 = \beta_i^2 \sigma_m^2 + \sigma^2(e_i)$$

The first term reflects *systematic risk* that impacts all securities (perhaps to different degrees due to differences in the betas) and the second term reflects firm-specific risk.

Covariance of Returns

Because a common factor, m, impacts the variance of returns for each security, the covariance of returns for any two securities is:

$$Cov(r_i, r_j) = \beta_i \beta_j \sigma_m^2$$

Single Index Model

One limitation of the single factor model that we just presented is that there is no insight into how to quantify the single factor, m. But suppose we made the assumption that the S&P 500 index could be used as a proxy for the single factor and then used this more practical model to make estimates of the betas, variances and covariances?

Here, we'll use M to denote the S&P 500 index value and for convenience reflect the *excess* returns over the risk free rate using $R = r - r_f$ instead of r. Then, we can depict the excess returns for any specific stock, as:

$$R_i = \alpha_i + \beta_i R_M + e_i$$

Notice that this formula indicates that the return in excess of the risk free rate is equal to a constant, denoted by alpha, plus a return due to movements in the overall market and beta plus an unanticipated firm-specific component, e_i.

Variance of Return

Using the formula shown above for the excess return over the risk free rate, we can write the variance of the excess return as:

$$\begin{aligned}\sigma_i^2 &= \text{Variance}(\alpha_i) + \text{Variance}(\beta_i R_M) + \text{Variance}(e_i)\\ &= \beta_i^2 \sigma_M^2 + \sigma^2(e_i)\end{aligned}$$

Covariances

The benefit of assuming an index model exists is to be able to estimate all of the covariances between risky assets for use in the portfolio optimization procedures we covered earlier. Here, it can be shown that the covariance between the returns of any two assets is:

$$Cov(R_i, R_j) = \beta_i \beta_j \sigma_M^2$$

This allows us to estimate the $N(N-1)/2$ covariances needed for the Markowitz model by simply estimating N betas and the variance of the market return. In addition, by not having to estimate covariances among wildly disparate types of firms, such as IBM and GM, market professionals are able to specialize in specific industries and know that covariances are driven by correlations with the overall market.

Diversification

The results above can be extended to portfolios of stocks. It turns out that if we have N stocks in a portfolio then the portfolio α is just the weighted average of the individual α's, the portfolio β is just the weighted average of the individual β's and the unique risk is just the weighted average of each of the firm-specific components (e_i).

More importantly, the variance of returns for the portfolio can be written as:

$$\sigma^2 = \beta_P^2 \sigma_M^2 + \sum w_i^2 \sigma^2(e_i)$$

This simply says that the portfolio risk is equal to a **systematic** component and the weighted sum of all the **unique** risks.

But similar to what we did in the previous chapter, as N gets large and if all $w_i = 1/N$, then the second term approaches zero and all that matters is the portfolio beta. The unique risk is diversified away and only the *systematic* risk matters.

Estimating Beta

To estimate beta we can use historical return data and the single index model of the form:

$$R_i = \alpha_i + \beta_i R_M + e_i$$

and then simply run a regression to estimate the slope of the regression line.

See the textbook for a review of key regression concepts that may be relevant when using an index model.

Portfolio Construction and the Index Model

A key advantage of the index model is the framework it creates for separating macroeconomic analysis (the analysis of economy-wide issues affecting the overall market) and security analysis (the analysis of the specific issues affecting individual stocks — the alphas).

By recognizing that the returns reflect both a common systematic component and a firm-specific component, an investment company can separate the estimation of the risk and risk premium for the overall market from the analysis of these firm-specific issues and ensure that the same macroeconomic analysis is used consistently.

Using the estimates of the risk and risk premiums for the market index, along with firm-specific beta estimates, we can identify the market-driven expected return and use it as a benchmark. Then security analysis can focus on trying to identify firms with a positive alpha — a return over and above that which can be expected based solely on its sensitivity to the market index and the expected risk premium for the market.

The Index Portfolio

In the absence of any firm-specific analysis, the market index (S&P 500 in this case) can be thought of as a passive portfolio against which various possible active portfolios can be judged against. These active portfolios would select weights on the stocks in the index that differ from their weights in the index in the hopes of improving the risk-return relationship.

Note that the passive portfolio would have the maximum diversification benefits of all possible portfolios. However, if through security analysis you could identify specific stocks with positive alphas, then it might make sense to depart from efficient diversification in favor of positive overall alpha.

The Optimal Risky Portfolio

As in the previous chapter, we could identify an optimal portfolio of risky assets by simply maximizing the Sharpe ratio. The only difference is that we would take advantage of the index model to help us quantify the alphas, betas and residual variances for each stock. The alphas and betas would help to quantify, along with the expected market risk premium, the expected returns for each security; the betas and the market variance would help to quantify the covariances; and the betas, residual variances and covariances would help to quantify the overall portfolio variance.

However, it turns out that there's an easier way to get to the same result (this approach is referred to as the Treynor-Black procedure). The idea is that the optimal portfolio will have a combination of w_A^* invested in an *active portfolio* and the rest invested in the index portfolio, a *passive portfolio*. To determine the weights of the stocks in the active portfolio and the weights between the active and passive portfolio, the following steps are followed:

1. Calculate an initial weight for each stock based on its ratio of alpha to residual variance. This gives more weight to the stocks whose non-market returns (the alpha) are largest relative to their non-market risk (the residual variance). Stocks with negative alphas will be given a negative weight (shorted).

2. Scale the weights described above so that they add to one.

3. Compute the alpha for the active portfolio using the scaled weights from (2) above.

4. Compute the residual variance of the active portfolio, denoted $\sigma^2(e_A)$, again using the weights from (2) above.

5. Calculate an initial weight for the active portfolio (which is accurate only if the beta for the active portfolio is 1.0) using the following ratio:

$$w_A^0 = \frac{\alpha_A/\sigma^2(e_A)}{E(R_M)/\sigma_M^2}$$

   This weight for the active portfolio reflects the non-market returns adjusted to reflect the non-market risk relative to the market returns adjusted to reflect the market risk.

6. Calculate the beta of the active portfolio using the weights from (2) above.

7. Adjust the weight in (5) above to account for the actual beta of the active portfolio, using the formula:

$$w_A^* = \frac{w_A^0}{1 + (1 - \beta_A)w_A^0}$$

8. The expected return and variance of the optimal risky portfolio can now be found. It will have $1 - w_A^*$ invested in the market portfolio and the balance invested in the active portfolio, in proportion to the weights in (2) above.

Now, let me go through the intuition behind these steps.

Suppose you had a selection of stocks for which you've been able to do security analysis and measure the alphas, the non-market driven component of their expected returns, as well as the residual risk, which is the non-market driven component of its total variability. The ratios of the alphas to residual variance for each stock would determine, on a relative basis, which ones to hold more or less of in your active portfolio. As noted above, stocks with higher alphas should get more weight, but stocks with more residual variance should get less weight. Once these relative weights are found (Step 1 above) we scale them so they add to 100% (Step 2 above). Using these weights, we can calculate the average active portfolio alpha and average residual variance (Steps 3 and 4).

Next, we know that this portfolio of selected stocks may not have adequate diversification on its own, so we want to split our investment between this active portfolio and a passive portfolio, such as an index. How would this be done? We'd want to give more weight (w_A) to the active portfolio the higher the portfolio alpha and the lower the portfolio residual variance, so some multiple of the ratio $\alpha_A/\sigma^2(e_A)$ is going to be helpful. Similarly, we want to give more weight to the passive portfolio the higher it's expected return and the lower its total risk, so a multiple of $E(R_M)/\sigma_M^2$ is going to be helpful. Taking the ratio of these two proportions, as shown in Step 5, will give us an initial weight, which it turns out is only appropriate if the beta of the active portfolio is equal to 1.0.

To adjust for the fact that the beta of the active portfolio may not be equal to 1.0, we recall that the whole rationale for adding a passive portfolio to the mix was to get more diversification. The higher the active portfolio beta is, the less diversification we are going to achieve and therefore the less we want to allocate to the passive portfolio. The adjustment in Step 7 accomplishes this, giving us the proportion in the active portfolio.

Along with the weights for this active portfolio found in Step 2 we can get the weight for each of the stocks. The balance of the weight is in the passive portfolio.

Note: You can download my edited version of the spreadsheet used in the BKM reading, which contains all of the calculations described above, from my website.

Information Ratio

When the index model is used as described above to identify the optimal portfolio, the Sharpe ratio of the optimal portfolio (S_P) will exceed the Sharpe ratio of the passive index portfolio (S_M) by an amount that depends on the ratio of the active portfolio's ratio of alpha to residual standard deviation, as shown below:

$$S_P^2 = S_M^2 + \left[\frac{\alpha_A}{\sigma(e_A)}\right]^2$$

The ratio of the alpha to the residual standard deviation is referred to as the *Information Ratio*. So another way to see the logic of the procedure outlined above is that its goal is to maximize the information ratio of the active portfolio, which occurs when each security's weight in the active portfolio is proportional to its own information ratio. The security's information ratio reflects the key trade-off associated with tilting the weights towards an individual asset — you gain from the security's alpha but add to the portfolio's residual risk.

Finally, this shows clearly that if the alpha's for all assets were zero, the optimal portfolio would have all of its weight on the passive portfolio. Thus, the goal of security analysis is to uncover positive alpha assets and use that insight to deviate from the passive market portfolio weights.

Comparing Index Model and Full-Covariance Model

Recall that the full-covariance (Markowitz) approach made use of the full matrix containing the covariances of each pair of assets. Making use of such information ought to allow us to identify more efficient portfolios. But keep in mind that this requires the estimation of thousands of values that are subject to significant estimation risk.

By contrast, the single index model makes a simplifying assumption that these covariances are driven by a single common factor and that the error terms are uncorrelated. While this results in a loss of some flexibility and loses some of the covariance that results from the possible correlation of the error terms, the practical advantages are significant.

In addition, as previously noted the index model allows for the separation of macroeconomic analysis and security analysis, allowing whatever benefits security analysis (the hunt for alpha) can achieve to be incorporated into the portfolio construction process.

Other Practical Aspects of Index Models

It was noted earlier that the single index model can be used to estimate alphas and betas using simple linear regression of the returns for a given stock against the returns for the index. However, there are a few complications.

Beta Books

Various firms regularly publish estimated betas for firms, however they often do a couple of things slightly differently than discussed above.

First, they use the total returns rather than excess returns. If the risk free rate is constant this produces the same beta estimates, but minor differences will arise in practice. In reality, the risk free rate is not constant but its variability is so small that its effect on beta is negligible. However, this small difference does cause their alpha parameter to be a bit different.

Second, they often use stock prices only and ignore dividends.

Adjusted Betas

Betas tend to regress towards 1.0 over time. This occurs for two reasons. First, firms are often formed to offer something fairly unique in terms of products or management, but as they grow they will tend to diversify and begin to look more like other firms, causing their betas to become more like the average firm's beta. Second, there are statistical issues associated with estimating beta (substantial estimation error) and it would make sense to use a Bayesian approach with a prior estimate of 1.0 (essentially a credibility adjustment). There are various ways to reflect this tendency to regress towards 1.0, but Merrill uses the following simple formula:

$$\text{Adjusted Beta} = \frac{2}{3}\,(\text{Estimated Beta}) + \frac{1}{3}\,(1.0)$$

Predicting Betas

Another issue with estimating beta is that we usually are interested in forecasting future betas. If we depict the current beta as a function of its past beta, then we might use a model like:

$$\text{Forecast Beta} = a + b(\text{Current Beta})$$

where the a and b parameters would be based on past relationships between betas and would reflect the tendency for betas to regress towards 1.0.

In addition, we could add other variables besides just past betas, such as firm size, debt ratio or even an industry factor.

Tracking Portfolios

Suppose you identify a portfolio that you believe earns a positive alpha, which represents a positive expected return over and above the risk free rate and the return based on the systematic risk (the beta) and the market return. You could create a tracking portfolio that consists of T-bills and the S&P 500 Index so that the tracking portfolio has the same systematic risk as the portfolio that you think has a positive alpha. By buying the portfolio and at the same time shorting the tracking portfolio, the systematic exposure to the market will be eliminated. Your returns will consist solely of the positive alpha and the risk from the non-systematic residual risk. With a well diversified portfolio, this residual risk should be small, allowing you to earn the positive alpha without any exposure to the overall market.

Practice Questions

Question 1. Assume a Single Index Model for stock returns applies and that for IBM stock the parameters were estimated as $\alpha = 1\%$ and $\beta = .9$. Also assume that the expected return for the market is 16%, the standard deviation of the market returns is 20%, the risk free rate is 5% and the standard deviation of the error terms in the single index model is 30%. What is the expected return and variance for IBM stock?

Solution. The single index model says that $R_i = \alpha_i + \beta_i R_M + e_i$ so that:

$$\begin{aligned} E(R_i) &= \alpha_i + \beta_i E(R_M) \\ &= 1\% + .9(16\% - 5\%) \\ &= 10.9\% \end{aligned}$$

This is the expected excess return over the risk free rate, hence the expected return is 15.9%.

We also know that the variance of this stock is:

$$\begin{aligned} \sigma_i^2 &= \beta_i^2 \sigma_M^2 + \sigma^2(e_i) \\ &= .9^2(.2^2) + .3^2 \\ &= .122 \end{aligned}$$

Question 2. Assume again that the single index model applies and that the β for Cisco's stock is 1.4. What is the covariance between IBM and Cisco's stock returns?

Solution. The text gives the formula $\sigma_{ij} = \beta_i \beta_j \sigma_M^2$ so the covariance between IBM and Cisco must be $\sigma_{IBM,CISCO} = (.9)(1.4)(.2^2) = .0504$.

Question 3. Assume you own a large, well diversified portfolio of stocks with an average beta of 1.1. If the standard deviation of the returns on the market is 30%, what is the variance of your portfolio?

Solution. For a portfolio,

$$\sigma_p^2 = \beta_p^2 \sigma_M^2 + \sum w_i^2 \sigma^2(e_i)$$

and when N is large, the second term approaches zero, so you can estimate the variance as simply $(1.1^2)(.3^2) = .1089$.

Question 4. Assume that for the past six months, the monthly returns in excess of the risk free T-Bill rate (in percentage terms) for Merck's stock and for the S&P 500 were as shown below. Estimate the parameters alpha and beta of a Single Index Model for Merck.

TABLE 1. Monthly Excess Returns

Month	Merck	S&P 500
Jan	6.00	8.00
Feb	-3.00	2.00
Mar	-8.00	1.00
Apr	-7.00	1.00
May	8.00	6.00
Jun	3.00	5.00
Average	-0.17	3.83

Solution. The formula for the estimator for β using standard linear regression is:

$$\hat{\beta} = \frac{\sum (R_i - \bar{R}_i)(R_m - \bar{R}_m)}{\sum (R_m - \bar{R}_m)^2}$$

The table below walks you through these calculations.

TABLE 2. Calculating Beta

$(R_i - \bar{R}_i)$	$(R_m - \bar{R}_m)$	$(R_i - \bar{R}_i)(R_m - \bar{R}_m)$	$(R_m - \bar{R}_m)^2$
6.17	4.17	25.69	17.36
-2.83	-1.83	5.19	3.36
-7.83	-2.83	22.19	8.03
-6.83	-2.83	19.36	8.03
8.17	2.17	17.69	4.69
3.17	1.17	3.69	1.36
	Sum	**93.83**	**42.83**

The beta estimate is then 93.83/42.83 = 2.19. The estimate for alpha is:

$$\begin{aligned}\alpha &= \text{Average } R_i - \beta(\text{Average } R_m)\\ &= -.17 - 2.19(3.83)\\ &= -8.6.\end{aligned}$$

Question 5. Assume that you are trying to estimate beta for GE's stock and you used data for the return on GE and the return on the market over the past 60 months to get a beta estimate of 1.5 with a standard error of .7. What estimate would Merrill Lynch use for their Adjusted Beta?

Solution. Adjusted Beta = .67(1.50) + .33(1) = 1.33.

Question 6. In the textbook the authors summarize an alternative to the full Markowitz procedure for determining the optimal risky portfolio, known as the Traynor-Black procedure, that makes use of the assumption that stock returns follow an index model. Their procedure finds the weights for a portfolio containing the market index as well as an active portfolio containing specific stocks. The active portfolio weights can be thought of as deviations (or tilts) away from a purely passive portfolio, with the weights driven largely by the analyst's assumptions regarding the alphas for the individual stocks as well as the residual (non-market risk).

Identify an advantage and disadvantage for each of these two approaches to portfolio construction.

Solution. The Markowitz approach allows you to take into account the full covariance matrix, but introduces potentially significant parameter estimation errors.

In contrast, the Traynor-Black approach ignores sources of covariance unrelated to the common index and may therefore miss opportunities for additional diversification. This is a disadvantage, but the Traynor-Black approach gives more insight into the separation of security analysis (the search for alpha) from macro- or market-wide forecasting.

Question 7. In the textbook the authors summarize an alternative to the full Markowitz procedure for determining the optimal risky portfolio that makes use of the assumption that stock returns follow an index model. Their procedure finds the weights for a portfolio containing the market index as well as an active portfolio containing specific stocks. The active portfolio weights can be thought of as deviations (or tilts) away from a purely passive portfolio, with the weights driven largely by the analyst's assumptions regarding the alphas for the individual stocks as well as the residual (non-market risk).

The authors' website contains a spreadsheet that implements this procedure using assumptions regarding the alphas and betas for several stocks as well as estimates of the standard deviations and residual errors from a single index model for the stocks as well as the market index. The following is a screen shot of their model:

FIGURE 1. Using Index Model to Form Optimal Portfolio

	A	B	C	D	E	F	G	H	I	J
1	**Panel 4: Macro Forecast and Forecasts of Alpha Values**									
2										
3										
4		S&P 500	*HP*	*DELL*	*WMT*	*TARGET*	*BP*	*SHELL*		
5	Alpha	0.0000	0.0150	-0.0100	-0.0050	0.0075	0.0120	0.0025		
6	Beta		2.0348	1.2315	0.6199	1.2672	0.4670	0.6736		
7	Risk premium	0.0600								
8										
9	**Panel 5: Computation of the Optimal Risky Portfolio**									
10										
11		**S&P 500**	***Active Portfolio A***	**HP**	**DELL**	**WMT**	**TARGET**	**BP**	**SHELL**	**Overall Portfolio**
12	$s^2(e)$			0.0705	0.0572	0.0309	0.0392	0.0297	0.0317	
13	$a/s^2(e)$		0.5505	0.2126	-0.1748	-0.1619	0.1911	0.4045	0.0789	
14	$W^0(i)$		1.0000	0.3863	-0.3176	-0.2941	0.3472	0.7349	0.1433	
15	$[W^0(i)]^2$			0.1492	0.1009	0.0865	0.1205	0.5400	0.0205	
16	a_A		0.0222							
17	$s^2(e_A)$		0.0404							
18	W_0		0.1691							
19	W^*	0.8282	0.1718							
20	Beta	1.0000	1.0922							1.0158
21	Risk premium	6.00%	8.78%							6.48%
22	SD	13.58%	24.97%							14.22%
23	Sharpe Ratio	0.4420	0.3514							0.4556

Without doing any calculations, describe which cells would change if the analyst changed his estimate of Dell's alpha from $-.01$ to $+.02$.

Solution. The point of this question is simply to make sure you understand the nature of the calculations rather than the specific formulas. What you should know is that because the Dell alpha changes, the presumption would be that the weight on Dell in the active portfolio would change (to be higher), which would initially affect cell E13. Then, the values in Row 14 would all change so that the weights in Row 13 are rebalanced to sum to 1.00.

Note that this would also cause the average alpha and the weighted standard errors in cells C16 and C17 to change in ways that you can't readily tell without doing calculations, but the end result will be a higher weight being given to the active portfolio and less weight given to the index. These weights are shown in cells B19 and C19 and depend upon an intuitive ratio of the average alphas scaled by the weighted standard errors to the average beta scaled by the market variance. That is, the more non-market return (alpha) that can be obtained, adjusted to reflect the non-market risk, relative to the amount of market-related return (beta), adjusted to reflect the market risk, the more you would want to tilt away from the passive portfolio and hold more of the active portfolio.

Of course, now that more weight is being given to Dell in the active portfolio, the active portfolio beta shown in cell C20 will be higher than before (Dell's beta was higher than the previous average beta). This will impact the risk premium on the active portfolio as shown in cell C21.

The end result is a higher expected risk premium and lower standard deviation, which produces a higher Sharpe ratio.

The values that result from this change in the Dell alpha are as follows:

FIGURE 2. Using Index Model to Form Optimal Portfolio

	A	B	C	D	E	F	G	H	I	J
1	**Panel 4: Macro Forecast and Forecasts of Alpha Values**									
2										
3										
4		S&P 500	*HP*	*DELL*	*WMT*	*TARGET*	*BP*	*SHELL*		
5	Alpha	0.0000	0.0150	0.0200	-0.0050	0.0075	0.0120	0.0025		
6	Beta		2.0348	1.2315	0.6199	1.2672	0.4670	0.6736		
7	Risk premium	0.0600								
8										
9	**Panel 5: Computation of the Optimal Risky Portfolio**									
10										
11		**S&P 500**	***Active Portfolio A***	**HP**	**DELL**	**WMT**	**TARGET**	**BP**	**SHELL**	**Overall Portfolio**
12	$s^2(e)$			0.0705	0.0572	0.0309	0.0392	0.0297	0.0317	
13	$a/s^2(e)$		1.0750	0.2126	0.3496	-0.1619	0.1911	0.4045	0.0789	
14	$W^0(i)$		1.0000	0.1978	0.3253	-0.1506	0.1778	0.3763	0.0734	
15	$[W^0(i)]^2$			0.0391	0.1058	0.0227	0.0316	0.1416	0.0054	
16	a_A		0.0163							
17	$s^2(e_A)$		0.0151							
18	W_0		0.3302							
19	W^*	0.6514	0.3486							
20	Beta	1.0000	1.1602							1.0559
21	Risk premium	6.00%	8.59%							6.90%
22	SD	13.58%	19.98%							14.96%
23	Sharpe Ratio	0.4420	0.4297							0.4613

Question 8. The table below contains the results, with some values missing, of the procedure discussed by the authors to determine the optimal portfolio when using a single index model:

FIGURE 3. Using Index Model to Form Optimal Portfolio

	A	B	C	D	E	F	G	H	I	J
1	**Panel 4: Macro Forecast and Forecasts of Alpha Values**									
2										
3										
4		S&P 500	*HP*	*DELL*	*WMT*	*TARGET*	*BP*	*SHELL*		
5	Alpha	0.0000	0.0150	-0.0200	-0.0050	0.0075	0.0120	0.0025		
6	Beta		2.0348	1.2315	0.6199	1.2672	0.4670	0.6736		
7	Risk premium	0.0600								
8										
9	**Panel 5: Computation of the Optimal Risky Portfolio**									
10										
11		**S&P 500**	***Active Portfolio A***	**HP**	**DELL**	**WMT**	**TARGET**	**BP**	**SHELL**	**Overall Portfolio**
12	$s^2(e)$			0.0705	0.0572	0.0309	0.0392	0.0297	0.0317	
13	$a/s^2(e)$		0.3757	0.2126	-0.3496	-0.1619	0.1911	0.4045	0.0789	
14	$W^0(i)$		1.0000	0.5660	-0.9307	-0.4309	0.5088	1.0769	0.2100	
15	$[W^0(i)]^2$			0.3204	0.8663	0.1857	0.2588	1.1596	0.0441	
16	a_A		0.0465							
17	$s^2(e_A)$		0.1238							
18	W_0		0.1154							
19	W^*	0.8842	0.1158							
20	Beta	1.0000	1.0274							1.0032
21	Risk premium	6.00%								
22	SD	13.58%	37.85%							
23	Sharpe Ratio	0.4420								

Assume that the variance of the market returns is .01843 and complete the missing values in cells C21, C23, J21, J22 and J23 of the table.

Solution. First, cell C21 reflects the expected risk premium on the active portfolio. It is the sum of the active portfolio average alpha and the product of the active portfolio beta and the market risk premium, or $.0465 + 1.0274(.06) = 10.82\%$.

Cell C23 is simply the ratio of the active portfolio risk premium and its standard deviation, .2857.

Cell J21 reflects the total portoflio risk premium, which uses the weights in cells B19 and C19 and the risk premiums in cells B21 and C21 to get an average risk premium of 6.56%.

Cell J22 is the standard deviation of the returns for the complete portfolio. The variance of returns is comprised of two elements: i) the squared value of the overall beta in J20 times the market variance (given as 0.01843) plus ii) the squared value of the active portfolio weight in cell C19 times the active portfolio squared standard errors in cell C17. The standard deviation is then the square root of this variance.

Finally, Cell J23 is just the ratio of the risk premium over the standard deviation, which in this case turns out to be .4613.

These values are all shown in the following completed table:

FIGURE 4. Using Index Model to Form Optimal Portfolio

	A	B	C	D	E	F	G	H	I	J
1	**Panel 4: Macro Forecast and Forecasts of Alpha Values**									
2										
3										
4		S&P 500	*HP*	*DELL*	*WMT*	*TARGET*	*BP*	*SHELL*		
5	Alpha	0.0000	0.0150	-0.0200	-0.0050	0.0075	0.0120	0.0025		
6	Beta		2.0348	1.2315	0.6199	1.2672	0.4670	0.6736		
7	Risk premium	0.0600								
8										
9	**Panel 5: Computation of the Optimal Risky Portfolio**									
10										
11		**S&P 500**	***Active Portfolio A***	**HP**	**DELL**	**WMT**	**TARGET**	**BP**	**SHELL**	**Overall Portfolio**
12	$s^2(e)$			0.0705	0.0572	0.0309	0.0392	0.0297	0.0317	
13	$a/s^2(e)$		0.3757	0.2126	-0.3496	-0.1619	0.1911	0.4045	0.0789	
14	$W^0(i)$		1.0000	0.5660	-0.9307	-0.4309	0.5088	1.0769	0.2100	
15	$[W^0(i)]^2$			0.3204	0.8663	0.1857	0.2588	1.1596	0.0441	
16	a_A		0.0465							
17	$s^2(e_A)$		0.1238							
18	W_0		0.1154							
19	W^*	0.8842	0.1158							
20	Beta	1.0000	1.0274							1.0032
21	Risk premium	6.00%	10.82%							6.56%
22	SD	13.58%	37.85%							14.22%
23	Sharpe Ratio	0.4420	0.2857							0.4613

Selected Old Exam Questions for Part 1

The following questions relevant for this section appeared on the Old CAS Exam 8 from 2000 to 2010 and on the CAS Exam 9 since 2011.

BKM 6	BKM 7	BKM 8
2003 Q14	2000 Q7	2001 Q7
2003 Q4	2000 Q8	2003 Q12
2004 Q3	2001 Q6	2003 Q7
2005 Q4	2002 Q7	2004 Q8
2006 Q4	2003 Q5	2005 Q8
2007 Q1	2005 Q5	2013 Q3
2009 Q1	2005 Q6	2014 Q4
2010 Q1	2006 Q5	2015 Q3
2013 Q1	2007 Q2	
2014 Q1	2008 Q1	
2014 Q5	2009 Q2	
2015 Q1a	2010 Q2	
	2010 Q3	
	2011 Q1	
	2012 Q1	
	2013 Q2	
	2015 Q1b	
	2015 Q1c	
	2015 Q2	

For some of these questions I have provided the text of the question and an explanatory solution. These were selected either because they are representative of the questions you are likely to be asked on future exams or because they contain an element that is particularly worthwhile to point out. For the other questions, the CAS solutions should be sufficient to confirm whether your answer is correct.

Important Note: The solutions shown here are intentionally detailed. They contain thorough explanations of the concepts and formulas used to reinforce the main points from the readings and provide an additional teaching opportunity. **Actual exam responses should be more concise than what is shown here, along the lines of what you will see in the solutions that the CAS releases.**

2007 Exam Question 1

You are given the following information:

- A risky portfolio has an expected return of 16% and a standard deviation of 25%
- The T-bill rate is 6%

a. Suppose you invest 60% of your funds in the risky portfolio and 40% in a T-bill money market fund. Calculate the expected value and the standard deviation of the rate of return of the portfolio.

The expected return is a weighted average of the expected returns of the components of the portfolio, so $E(r) = .6(16\%) + .4(6\%) = 12\%$.

The standard deviation of a portfolio depends on the weights, the standard deviations and the covariance of the components. When one of the portfolios is risk free, its standard deviation is zero and its covariance with the other components is zero as well. As a result, the formula is simplified as the weight on the risky portfolio times its standard deviation, $\sigma = .6(25\%) = 15\%$.

b. Determine the equation of the Capital Allocation Line (CAL) of the risky portfolio and graph the CAL. Plot the position of the overall portfolio on the CAL graph. Label all items properly.

The CAL depicts the relationship between the expected return and standard deviation of a portfolio that is allocated $y\%$ to a risky portfolio and $(1 - y\%)$ to a risk free portfolio. In Part (a) above this allocation was 60%/40% and resulted in an expected return of 12% and a standard deviation of 15%. If the allocation were 100%/0% the expected return would be 16% and the standard deviation 25%. Or, if the allocation were 0%/100% the expected return would be 6% and the standard deviation 0%. The CAL is the line that connects all of these points.

To write the equation of the line, we can simply use any of these two points and basic algebra. For instance, we could use the slope-intercept equation and note that the intercept is 6% (that's when the standard deviation is zero) and the slope is easy to find using the 100%/0% allocation relative to this intercept, since the rise is 16%-6% = 10% and the run is 25% - 0% = 25%.

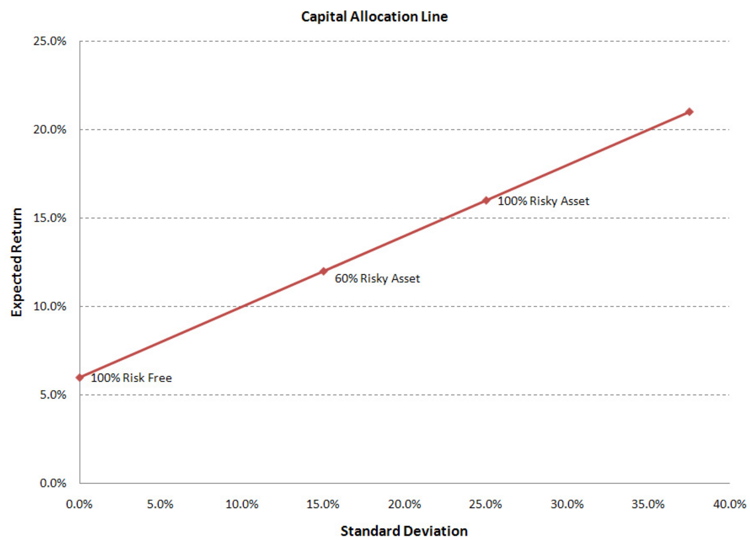
The equation of this line is $E(r) = 6\% + (10\%/25\%)\sigma$. This can be written as $E(r) = 6\% + .4\sigma$.

Note that the generic version of this given in the text is:

$$E(r_c) = r_f + \frac{\sigma_c}{\sigma_p}[E(r_p) - r_f]$$

We can graph this as follows:

FIGURE 1. Capital Allocation Line



2007 Exam Question 2

Available securities include two risky stock funds, X and Y, and T-bills. Data for the securities follows:

	Exp Return	Std. Deviation
Fund X	8.00%	22.00%
Fund Y	27.00%	70.00%
T-Bills	6.00%	0.00%

The correlation coefficient between Funds X and Y is -0.15.

a. Determine the expected return and standard deviation of the optimal risky portfolio.

The optimal risky portfolio is the combination of Funds X and Y that has the highest Sharpe Ratio, which is the ratio of the expected return in excess of the risk free rate over the standard deviation. The first thing we need to do is determine the composition of this portfolio (how much is in X and Y) and then we can easily find the expected return and standard deviation.

In the notes we derived the solution to this problem for the special case with only two risky assets and found that the optimal weight on Fund X could be written as:

$$
\begin{aligned}
w_X &= \frac{[E(r_X) - r_f]\sigma_Y^2 - [E(r_Y) - r_f]\sigma_{XY}}{[E(r_X) - r_f]\sigma_Y^2 + [E(r_Y) - r_f]\sigma_X^2 - [E(r_X) - r_f + E(r_Y) - r_f]\sigma_{XY}} \\
&= \frac{(.02)(.7^2) - .21(.22)(.7)(-0.15)}{(.02)(.7^2) + (.21)(.22^2) - [.02 + .21](.22)(.7)(-0.15)} \\
&= 0.579
\end{aligned}
$$

Note that in the question you were given the correlation coefficient, ρ, but in the formula shown above we need the covariance, denoted σ_{XY}. The relationship between the two is $\sigma_{XY} = \sigma_X \sigma_Y \rho$.

Now that we know the composition, we can find the expected return and standard deviation using the basic equations from the text.

$$
E(r) = (8\%)w_X + (27\%)(1 - w_X) = 15.98\%
$$

The variance of his portfolio will be:

$$
\begin{aligned}
\sigma^2 &= w_X^2(22\%)^2 + (1 - w_X)^2(70\%)^2 + 2w_X(1 - w_X)(22\%)(70\%)(-0.15) \\
&= 0.0916
\end{aligned}
$$

And then the standard deviation will be $\sigma = 30.26\%$.

b. "A" is the index of an investor's risk aversion. Calculate how much an investor with a risk aversion index of 4 will invest in each of Funds X and Y and in T-Bills. Assume that the investor's utility function is given by the equation:

$$U = E(r) - .5A\sigma^2$$

Note that the original question did not specify the utility function. In addition, note that the new edition of the textbook defines the utility function using different units than in earlier editions, so the sample solution will be different from the one shown below.

The task here is to first determine the optimal allocation to the optimal risky portfolio from the previous question and to the risk free asset for this particular investor. We can do this by writing the expected return and standard deviation of their complete portfolio (using the risk free rate and the results from part (a) above) for different weights, y, on the risky portfolio, plugging those into the utility function and maximizing the value of this function with respect to y.

The resulting formula for this will be:

$$\begin{aligned} y^* &= \frac{[E(r_p) - r_f]}{A\sigma_p^2} \\ &= \frac{.1598 - .06}{4(.3026^2)} \\ &= .272 \end{aligned}$$

This means that 27.2% will be invested in the risky portfolio and 72.8% will be invested in the risk free portfolio. Note that the risky portfolio contains 57.9% in X and 42.1% in Y, so the overall allocation is:

TABLE 1. Overall Allocation

	Allocation
Fund X	15.80%
Fund Y	11.50%
T-Bills	72.70%

Note that the 2008 Exam Question 1 was identical to this question, but with different numbers and one extra part about rebalancing the portfolio to maintain optimal proportions.

2003 Exam Question 4

Answer the questions below based on the following information about a risky portfolio that you manage, and a risk free asset:

- $E(r_p) = 11\%$
- $\sigma_p = 15\%$
- $r_f = 5\%$

a. Client A wants to invest a proportion of her total investment budget in your risky fund to provide an expected rate of return on her overall or complete portfolio equal to 8%. What will be the standard deviation of the rate of return on her portfolio?

When combining a risky portfolio with the risk free asset, with y representing the proportion invested in the risky asset, the expected return on the complete portfolio is given as:

$$E(r_c) = yE(r_p) + (1 - y)r_f$$

Since the risk free asset has a standard deviation of zero, the standard deviation of the complete portfolio is given as $\sigma_c = y\sigma_p$.

In this question, you need to first determine y before you can solve for the standard deviation of the complete portfolio. They gave you the criteria that the investor wanted the expected return to equal 8%, so $E(r_c) = 8 = y(11) + (1 - y)(5)$ which can be easily solved for $y = 50\%$.

Then the standard deviation is given by $\sigma_c = y\sigma_p = .5(15\%) = 7.5\%$.

b. Client B wants the highest possible return subject to the constraint that you limit his standard deviation to be no more than 12%. Which client is more risk averse? Explain why.

It is not entirely clear how complicated they intended to make this question. For an easy answer, note that if we assume that the two investors have selected their constraints so as to achieve their optimal portfolio holdings, then all we need to notice is that A's optimal portfolio has 7.5% standard deviation and B's optimal portfolio has 12% standard deviation. Since A selects a less risky portfolio, she must be more risk averse.

2001 Exam Question 6

You are given the following information:

Number of Securities	Portfolio Variance
1	32.100
2	21.180
20	11.352
500	10.304

- Each security has equal weight in a given portfolio.
- The variance of all securities is the same.
- The covariance between each pair of securities is the same.

a. Calculate the minimum variance possible for a portfolio constructed from this group of securities.

The formula for portfolio variance is:

$$\sigma_p^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} w_i w_j \sigma_{ij}$$

In the case where the weights are all the same (i.e. they are all $1/N$ when there are N securities) and the variances and covariances are all the same, then this is going to equal:

$$\sigma_p^2 = \frac{1}{N}\sigma^2 + \frac{N-1}{N}\sigma_{ij}$$

Further, since the question told you that $\sigma^2 = 32.1$, we can calculate the covariance between any pair, σ_{ij}, using the two asset case $21.18 = (1/2)(32.1) + (1 - 1/2)\sigma_{ij}$. Solving this we get $\sigma_{ij} = 10.26$.

From this, we see that the minimum possible variance as N gets very large is $\sigma_{ij} = 10.26$.

b. How many securities are needed to achieve an average portfolio variance that is within 5% of this minimum?

To get within 5% of this minimum, we need $\sigma_p^2 = 10.26(1.05) = 10.773$.

Now just solve for N such that:

$$\begin{aligned} 10.773 &= \frac{1}{N}\sigma^2 + \frac{N-1}{N}\sigma_{ij} \\ &= \frac{1}{N}(32.1) + \frac{N-1}{N}(10.26) \end{aligned}$$

Solving for N gives us $N = 42.57$ and so you need 43 securities.

2003 Exam Question 12

Assume that the single-index model for Stocks A and B is estimated with the following results:

- $R_A = 1.0\% + 0.95R_M + e_A$
- $R_B = -2.0\% + 1.2R_M + e_B$
- $\sigma_m = 25\%$
- $\sigma(e_A) = 32\%$
- $\sigma(e_B) = 8\%$

a. Find the standard deviation of each stock.

The single index model says that:

$$R_i = \alpha_i + \beta_i R_m + e_i$$

Taking the variance of this we get:

$$\sigma_A^2 = \beta_A^2 \sigma_m^2 + \sigma(e_A)^2 = .95^2(.25^2) + (.32^2) = .1588$$

And from this, $\sigma_A = 39.85\%$.

Similarly, for Stock B,

$$\sigma_B^2 = \beta_B^2 \sigma_m^2 + \sigma(e_B)^2 = 1.2^2(.25^2) + (.08^2) = .0964$$

And then $\sigma_B = 31\%$.

b. Find the covariance between stocks A and B.

If a single index model describes asset returns, then the covariance of any two assets is given by:

$$\sigma_{ij} = \beta_i \beta_j \sigma_m^2$$

So the covariance between these two stocks is:

$$\sigma_{ij} = (.95)(1.2)(.25^2) = .07125$$

2010 Exam Question 2

- The return of a risk-free asset is 5%
- An investment company offers a risky asset, with a Sharpe ratio of 0.20
- An investor wants to hold a portfolio consisting of the risky asset and the risk-free asset

a. Calculate the expected return of the portfolio if the investor wants the standard deviation of the portfolio to be 15%.

This is a typical question for this section of the syllabus — trivially easy formulas that need to be solved using algebra. The problem with these types of questions is that they often result in either noticing the right algebra steps right away and answering it in no time or taking the wrong path and spending an inordinate amount of time on it.

The quickest answer is to recognize that when you combine a risky asset and a risk-free asset in a portfolio the Sharpe ratio of the portfolio matches the Sharpe ratio of the risky asset. Therefore, we know that:

$$S = 0.20 = \frac{E(r_c) - r_f}{\sigma_c}$$

Since we are told that $\sigma_c = 15\%$, we know that:

$$E(r_c) = .20(.15) + .05 = 8\%$$

An even quicker way to get this was to recall that there was a specific formula for the expected return for any portfolio on the CAL:

$$\begin{aligned} E(r_c) &= r_f + \frac{\sigma_c}{\sigma_p}[E(r_p) - r_f] \\ &= r_f + \sigma_c \left[\frac{E(r_p) - r_f}{\sigma_p} \right] \\ &= .05 + .15 * (.2) \\ &= 8\% \end{aligned}$$

b. Graph the capital allocation line (CAL) associated with this portfolio. Plot the position of the overall portfolio on the CAL graph. Clearly label the axes, the CAL, the risk-free asset and the overall portfolio.

I won't show the actual graph (see other similar questions), but just note that you are not given any information with regard to the risky asset so you cannot locate that exactly on the graph. You just know it is some point on the CAL.

But as discussed above, the Sharpe ratio is constant along the CAL and identical to the Sharpe ratio for the risky asset. We can easily confirm this, $S = (.08 - .05)/.15 = .20$. So it is easy to identify the portfolio and the risk-free asset and then we know that the CAL is just the line through those points. The risky asset is along that line somewhere.

2012 Exam Question 1

There are three risky assets (A, B and C) and one risk-free asset (D). You are given the following information:

- Expected returns are: $r_A = 12\%, r_B = 10\%, r_C = 6\%$ and $r_D = 3\%$
- The coefficient of risk aversion is $A = 3$ for the investor with a utility function $U = E(r) - .5A\sigma^2$
- The weight given to Asset A in the optimal risky portfolio is 40%
- The standard deviation of the optimal risky portfolio is 16%
- the slope of the capital allocation line is 0.35

a. Constuct an optimal investment plan using some of each asset (A, B, C and D) and justify your proposed plan. Include brief descriptions of the proportion to be invested in each asset and the overall expected risk and return results.

This is largely an algebra problem that is made easier when you make use of the information you are given. We ultimately want to use the utility function to identify the optimal portfolio, but first we need to find the optimal risky portfolio.

We are told the slope of the CAL is 0.35, which tells us that the Sharpe ratio of the optimal risky portfolio is $S = 0.35$. We are also told that the standard deviation of the optimal risky portfolio is 16%, so that allows us to find the expected return of the optimal risky portfolio:

$$S = 0.35 = \frac{E(r) - r_f}{\sigma} = \frac{E(r) - 3\%}{16\%} \Rightarrow E(r) = 8.6\%$$

Now we need to solve for the three weights in the optimal risky portfolio that produces a portfolio with an expected return of 8.6%. This would be impossible with three unknowns, but we are told that the weight on A is 40%, so there is really only one unknown, w_B. The final weight is just $w_C = 1 - .4 - w_B$. The equation is:

$$8.6\% = (.4)(.12) + w_B(.10) + (1 - .4 - w_B)(.06)$$

From this, we can solve for $w_B = .05$ and $w_C = .55$.

Given what we know about the optimal risky portfolio, now we just need to find the weighting between this risky portfolio with $E(r) = 8.6\%$ and $\sigma = 16\%$ and the risk-free asset. Given the utility function, we can maximize the utility for different weights on the risky portfolio, y, and solve for the optimal weight:

$$y^* = \frac{[E(r_p) - r_f]}{A\sigma_p^2} = \frac{.086 - .03}{3(.16^2)} = 0.729$$

The final portfolio weights will then be as follows:

$$w_A = 0.729 * .40 = 0.2916$$

$$w_B = 0.729 * .05 = 0.03645$$

$$w_C = 0.729 * .55 = .40095$$

$$w_D = 1 - 0.729 = 0.271$$

These weights maximize the utility function for the investor in the most efficient way possible, combining the optimal risky portfolio with the risk-free asset.

The final expected return is 7.0824% and the final standard deviation is 11.664%.

Explain why the proportion invested in Asset D may be the only difference between the plans for investors with the same assets.

Investors with different levels of risk aversion may prefer higher- or lower-risk complete portfolios, but they will achieve this by selecting the same optimal risky portfolio and merely adjust the proportion invested in the risk-free asset, thus moving along the CAL.

2013 Exam Question 3

Given the following information:

- the expected risk premium of the market index is 8%
- the standard deviation of the market index is 20%
- A single-factor model is used to construct an optimal risky portfolio with two components:
  - a passive portfolio containing the market index
  - an active portfolio containing some combination of three stocks with the following estimated parameters:

Stock	α	$\sigma^2(e)$	β
1	0.06	0.16	0.80
2	0.10	0.20	1.00
3	0.15	0.24	1.20

Calculate the portion of the optimal risky portfolio that is invested in the active portfolio.

Following the steps outlined in the notes:

1. Calculate an initial weight for each stock based on its ratio of alpha to residual variance.

$$w_1 = .06/.16 = .375$$

$$w_2 = .10/.20 = .5$$

$$w_3 = .15/.24 = .625$$

2. Scale the weights described above so that they add to one.

$$w_1 = .25$$

$$w_2 = .33$$

$$w_3 = .417$$

3. Compute the alpha for the active portfolio using the scaled weights from (2) above.

$$\alpha_A = .25(.06) + .33(.10) + .417(.15) = .1108$$

4. Compute the residual variance of the active portfolio, again using the weights from (2) above.

$$\sigma^2(e_A) = .25^2(.16) + .33^2(.20) + .417^2(.24) = .0739$$

5. Calculate an initial weight for the active portfolio (which is accurate only if the beta for the active portfolio is 1.0) using the following ratio:

$$\begin{aligned} w_A^0 &= \frac{\alpha_A/\sigma^2(e_A)}{E(R_M)/\sigma_M^2} \\ &= \frac{.1108/.0739}{.08/(.2^2)} \\ &= .75 \end{aligned}$$

6. Calculate the beta of the active portfolio using the weights from (2) above.

$$\beta_A = .25(.8) + .33(1) + .417(1.2) = 1.033$$

7. Adjust the weight in (5) above to account for the actual beta of the active portfolio, using the formula:

$$\begin{aligned} w_A^* &= \frac{w_A^0}{1 + (1 - \beta_A)w_A^0} \\ &= \frac{.75}{1 + (1 - 1.033)(.75)} \\ &= .769 \end{aligned}$$

2015 Exam Question 1b

You are given the information below describing a three-security financial market, one rational investor's complete portfolio:

Security	Expected Return	Std Deviation	Portfolio Allocation
X	4%	0%	43%
Y	10%	20%	30%
Z	15%	30%	27%

Investors' utility function is given by $U = E(r) - .5A\sigma^2$.

Calculate the correlation coefficient of the returns of securities Y and Z.

We know from the portfolio allocation that the investor's risky portfolio consists of the following weights on securities Y and Z:

$$w_Y = \frac{30\%}{1 - 43\%} = 52.63\%$$

$$w_Y = \frac{27\%}{1 - 43\%} = 47.37\%$$

If the investor has selected his optimal portfolio, then the weight on security Y should have been:

$$w_Y = \frac{[E(r_Y) - r_f]\sigma_Z^2 - [E(r_Z) - r_f]\sigma_{YZ}}{[E(r_Y) - r_f]\sigma_Z^2 + [E(r_Z) - r_f]\sigma_Y^2 - [E(r_Y) - r_f + E(r_Z) - r_f]\sigma_{YZ}}$$

We can then use this equation and the information given to solve for $\sigma_{YZ} = 1.18\%$. This is the covariance, so to get the correlation, we just divide by the standard deviations and get $\rho_{YZ} = 19.66\%$.

Part 2

Equilibrium in Capital Markets

Part 2 Table of Contents

BKM Chapter 9: The Capital Asset Pricing Model

This chapter essentially continues the discussion of Portfolio Theory but adds assumptions regarding how all investors behave — essentially assuming that they all do the same sort of analysis that we just did. The result will be that the expected returns on risky assets will be a linear function of the expected return on the market portfolio.

All Investors Will Own the Market Portfolio

Suppose that all investors are rational, mean-variance optimizers who follow the Markowitz procedure discussed in the previous chapter, that they all use the same assumptions regarding expected returns and standard deviations, that they can buy or sell any asset and that they can borrow or lend at a risk-free rate of interest.

One immediate implication of this is that, because every investor makes the same assumptions and performs the same analysis, they will ultimately own the same portfolio of risky assets. This does not mean that they all choose the same mix of the risky asset and the risk free asset — that will depend on each person's risk aversion — but the point on the efficient frontier that determines each investor's Capital Allocation Line will be identical. And if that is the case, then this portfolio must be the Market Portfolio consisting of all shares in proportion to their market value.

Why must this be the market portfolio?

Assume that one stock, such as Delta Airlines, was not in the optimal portfolio that everyone wanted to own. With nobody owning Delta, its price will fall, forcing each investor to reevaluate the expected return for Delta and therefore causing the optimal mix to change. At some point, Delta will become desirable and then again, everyone will own each asset in proportion to its market value.

Passive Strategy is Efficient

We just saw that the optimal risky portfolio for investors is the market portfolio that contains all risky assets in proportion to their market values. This means that their Capital Allocation Line is actually the Capital Market Line we discussed in Chapter 8. This suggests that nobody actually has to do any security analysis — they can simply own an index fund of all risky assets.

Risk Premium of the Market Portfolio

Remember that after investors determine their optimal risky asset portfolio composition (in this case the Market Portfolio) their next task is to determine their preferred mix of the risk free asset and this risky asset.

From earlier chapters, we saw their allocation to the risky asset, y^*, depended on their risk aversion and equals:

$$y^* = \frac{[E(r_M) - r_f]}{A\sigma_M^2}$$

Since all borrowing and lending has to offset in the aggregate, on average $y = 1$. And therefore, on average:

$$E(r_M) - r_f = \bar{A}\sigma_M^2$$

Expected Returns on Individual Securities

There is a rather subtle but important point about what we have done so far that is easy to overlook. Note that given the assumptions we have made, we have concluded that each investor would draw an identical capital allocation line and that since the optimal risky portfolio is the Market Portfolio, we called this line the Capital Market Line (CML). This line tells us the risk and return trade-offs from investments that combine this portfolio with the risk free asset.

Note that there is no real reason why investors have to choose complete portfolios along this line — they are free to select any portfolio they like. For instance, they could invest 100% of their money in a single stock, which most likely lies at a point on the interior of the efficient frontier. However, only portfolio choices along the CML are *efficient*. So if we assume that investors try to optimize their selection, then we can conclude that they would in fact select portfolios along the line. And therefore the CML tells us the risk and return for any *efficient* portfolio.

But what does this tell us about the risk and return for an individual stock? Well, nothing at all. We can conclude that portfolios that are owned by investors lie along this line, but we know that no investors will simply own a single risky asset on its own, so we have no reason to believe that it lies on the CML.

Before getting to the textbook's explanation of how to answer this question about the risk and return of individual stocks, let's consider a different explanation that may be easier to understand. Then we will come back to the book's approach:

Deriving CAPM, Simple Method

From the above arguments, we know that any complete portfolio that investors would be willing to own (i.e. efficient portfolios) must have an expected return along the line connecting the risk free rate and the market portfolio—the Capital Market Line. Using basic point slope form, this line can be written as:

$$E(r_c) = r_f + \frac{E(r_M) - r_f}{\sigma_M}\sigma_c$$

Where $E(r_M)$ is the expected return on the market and σ_M is the standard deviation of the market portfolio.

This can also be written as:

$$E(r) = r_f + \text{Market Reward to Risk Ratio} * \text{Amount of Risk}$$

where the market reward to risk ratio is given by the equation:

$$\text{Market Reward to Risk Ratio} = \frac{E(r_M) - r_f}{\sigma_M}$$

and σ_c is the *amount of risk* in the complete portfolio.

Note that the formula for the expected return as a linear function of the market price of risk and the asset's standard deviation does NOT apply to all stocks or portfolios. It only applies to the efficient portfolios that lie along the Capital Market Line. It would be helpful to extend this analysis to all stocks. So why not just substitute σ_i for the stock in question?

We cannot do this because for individual stocks the total standard deviation (σ_i) is not a good measure of risk. For investors with large portfolios, their risk depends on the covariances, or the systematic risk, not the total risk of the stocks in the portfolio. Therefore, what we would prefer to use is a measure of the *marginal* risk that any one stock would add to our existing portfolio. This means we need to find the derivative of the portfolio risk with respect to the weight we put on stock i.

Remember that if we already assume that investors own the market portfolio, then their total portfolio variance can be written as:

$$\sigma_M^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} w_i w_j \sigma_{ij}$$

where the sums are taken over all N assets in the market portfolio and i not equal to j.

What is the marginal increase in risk from a small change in the weight for security i? Defining risk as the standard deviation of returns, then the marginal risk is given as follows:

$$\begin{aligned}
\text{Marginal Risk} &= \frac{\partial [\sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} w_i w_j \sigma_{ij}]^{1/2}}{\partial w_i} \\
&= \frac{1}{2} [\sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} w_i w_j \sigma_{ij}]^{-1/2} (2 w_i \sigma_i^2 + 2 \sum_{j \neq i} w_j \sigma_{ij}) \\
&= \frac{w_i \sigma_i^2 + \sum_{j \neq i} w_j \sigma_{ij}}{\sigma_M}
\end{aligned}$$

Notice that σ_{ij} is the covariance of stock i with stock j and we can think of σ_i^2 as the covariance of stock i with itself. Therefore, the numerator of the above equation is simply the weighted average of the covariance of stock i with each of the components of the market portfolio. Therefore, this numerator can be viewed quite simply as the covariance of stock i with the market portfolio, or

$$\text{Marginal Risk} = \frac{Cov(r_i, r_M)}{\sigma_M}$$

This should be a more appropriate risk measure for an individual stock, so we can plug that in as the amount of risk in the Capital Market Line equation to get:

$$E(r_i) = r_f + \frac{E(r_M) - r_f}{\sigma_M} \frac{Cov(r_i, r_M)}{\sigma_M}$$

Defining beta as follows:

$$\beta = \frac{Cov(r_i, r_M)}{\sigma_M^2}$$

we can rewrite the previous equation for the expected return on the stock as:

$$E(r_i) = r_f + \beta[E(r_M) - r_f]$$

This is known as the **Capital Asset Pricing Model.**

It simply says that the expected return for any stock is equal to the risk free rate plus a risk premium equal to beta times the difference between the expected return on the market and the risk free rate.

Notice that while the earlier analysis was done in terms of the standard deviation risk measure σ, CAPM allows us to think of everything in terms of β, which will now serve as our measure of risk for individual assets.

Derivation of CAPM, Textbook Presentation

The textbook's derivation of CAPM is similar to what we have just done, but they explain the steps differently. I will walk you through their steps and define the terms the way they do.

Recall that we began our discussion of portfolio theory by assuming that investors make portfolio choices in the following manner. First, they evaluate all possible combinations of risky assets to determine an optimal risky portfolio. When selecting this optimal risky portfolio, they seek to maximize their return and minimize their risk, using the variance of returns as their measure of risk. When we note that the investor can either borrow or lend at the risk free rate, there is one particular risky asset portfolio that is optimal for them.

Next, they combine this optimal risky portfolio with either borrowing at the risk free rate to increase (leverage) their overall risk and return or with lending at the risk free rate to decrease their overall risk and return. Their final portfolio composition will depend upon their attitudes towards risk, expressed in terms of their utility function, in order to maximize their expected utility.

If we assume that all investors have the same information regarding the expected returns and standard deviations of returns for all risky assets, and that they all follow this same procedure, then all investors will wind up selecting the *same* risky asset portfolio. However, if they each have their own unique attitudes regarding risk, they may still select different combinations of this risky portfolio and the risk free asset to form their complete portfolio.

In equilibrium, each investor's risky asset portfolio will have to mirror the *market portfolio*. In other words, if everyone chooses the same portfolio of risky assets, this portfolio has to contain risky assets in the same proportion as these assets exist in the total market portfolio. Similarly, the aggregate borrowing and lending will have to be equal.

To determine what this implies about the expected returns and risk of an individual security, consider the contribution to both the portfolio return and risk from any specific security. That is, how much do the portfolio return and risk change when the weight on a specific security is increased?

To keep things simple, consider the case for the investor who has y percent of his assets invested in the market portfolio (the optimal risky portfolio) and the rest invested in the risk-free asset. Then assume that the investor wants to increase the portfolio risk by adding some small amount, δ_i, of one particular risky asset to the portfolio. So that the investor doesn't have to modify his existing portfolio, assume that the amount to be invested is obtained by borrowing at the risk free rate.

The contribution that this change will make to the portfolio's risk premium is reflected in the additional return $\delta_i E(r_i)$ generated by the stock less the borrowing costs $\delta_i r_f$:

$$\text{Contribution to Risk Premium} = \delta_i[E(r_i) - r_f]$$

The contribution this change will make to the total portfolio risk is a bit trickier, but it can be shown to equal:

$$\text{Contribution to Variance} = 2\delta_i y Cov(r_i, r_M)$$

Taking the ratio of these two marginal amounts, we can define a **reward-to-risk ratio** for any particular stock as:

$$\frac{\text{Contribution to Risk Premium}}{\text{Contribution to Variance}} = \frac{\delta_i[E(r_i) - r_f]}{2\delta_i y Cov(r_i, r_M)} = \frac{E(r_i) - r_f}{2y Cov(r_i, r_M)}$$

That ratio shows how much the risk premium increases relative to the increase in risk when a small amount of one particular asset is added to the market portfolio.

Important Mathematical Note: The above explanation summarizes the main point of the textbook's explanation but does it slightly differently to avoid some confusing simplifications in the textbook. To see a more rigorous derivation of the calculation of the contribution to portfolio variance, you can refer to the appendix to the notes for this chapter immediately following the questions below.

An alternative way for an investor to increase risk is to simply increase the proportion, y, invested in the market portfolio.

Since the portfolio return is given by the formula:

$$E(r_c) = yE(r_M) + (1 - y)r_f$$

the marginal risk premium for a small change in y is equal to:

$$\text{Contribution to Risk Premium} = E(r_M) - r_f$$

And since the portfolio risk is equal to:

$$\sigma_c^2 = y^2\sigma_M^2$$

the marginal risk for a small change in y is equal to:

$$\text{Contribution to Variance} = 2y\sigma_M^2$$

This would produce a reward to risk ratio of:

$$\frac{\text{Contribution to Risk Premium}}{\text{Contribution to Variance}} = \frac{E(r_M) - r_f}{2y\sigma_M^2}$$

Given these two choices of either:

a) investing in a particular stock, or

b) investing in the overall market with varying levels of leverage,

investors have two possible measures of their expected rewards per unit of risk. Unless these two ratios are equal, investors will have a preference for one over the other and then natural market forces will cause the prices to adjust, the expected returns to change and the difference to disappear. In equilibrium, the reward-to-risk ratios will have to be equal.

This will lead to the following equality in equilibrium:

$$\frac{E(r_i) - r_f}{2yCov(r_i, r_M)} = \frac{E(r_M) - r_f}{2y\sigma_M^2}$$

Rearranging terms, this can be written as,

$$E(r_i) = r_f + [E(r_M) - r_f]\frac{Cov(r_i, r_M)}{\sigma_M^2}$$

Defining beta as follows:

$$\beta = \frac{Cov(r_i, r_M)}{\sigma_M^2}$$

we can rewrite the previous equation for the expected return on the stock as:

$$E(r_i) = r_f + \beta[E(r_M) - r_f]$$

Security Market Line

Recall that in earlier discussions we developed the Capital Market Line, which showed the risk and return trade-offs for efficient portfolios and defined risk in terms of standard deviation. We could similarly graph the CAPM equation. But note that the CAPM differs from the CML in two ways:

- it measures risk in terms of beta, the marginal increase in the portfolio risk, and

- it applies to all risky assets, not just efficient portfolios.

The graph of the CAPM equation is called the Security Market Line (SML).

The SML is important for portfolio management because it serves as a benchmark to judge whether a security's expected return is sufficient to compensate investors for the risks they assume, where risk is measured using beta.

Alpha

If a stock's expected return-beta relationship does not lie along the SML, then we say that it has a non-zero **alpha**, which is the difference between the fair and actual expected returns. If assets are priced fairly, then they will lie along the SML. If their prices are too high, their expected returns will be too low and will lie below the line; if their prices are too low then they will lie above the SML. This is an important perspective since it suggests that *the goal of portfolio management should be to identify assets with positive alpha*.

Tales from the FAR Side

The Words From the Street Box provides a review of the key elements of CAPM and then discusses the fact that recent empirical evidence (discussed in much more detail in Chapters 11-13) has found that beta may not actually explain returns as well as it was once thought. More importantly, there seems to be another measure that does explain returns — the ratio of a firm's book value relative to its market value.

Since firms with high book to market ratios seem to earn higher returns, this ratio might represent an additional risk factor not reflected in the standard CAPM. Perhaps the CAPM formula for the expected return should be adjusted by an additional risk premium, producing a *New Estimator of Expected Return* or NEER.

But what if, instead of the book to market ratio representing a source of additional risk, the higher returns for high book to market ratio firms is simply the result of investor mistakes that are eventually corrected? In other words, what if the high return for stocks with high book to market ratios is simply due to their price being too low (by mistake) initially and then the return is high because the mistake eventually gets corrected?

This will create a paradox. If investors are rational, then book to market ratios must in fact represent a risk that managers should take into account when determining the rates to use to discount project cash flows. If investors are irrational and are making mistakes like the one noted above, then the managers have to make a choice. If they want to worry about their current stock price, then they have to follow investors' irrational model and use the NEER estimates as their hurdle rates. But if they really care about the long term value of the firm, they should ignore the investors' errors and use the original CAPM, which Stein refers to as the *Fundamental Asset Risk* model, or FAR.

CAPM and the Single-Index Model

So far, we derived CAPM by first establishing that the market portfolio is efficient (it is the one portfolio that everyone owns) and then evaluated risk relative to this portfolio. This resulted in the risk premium for an asset being proportional to its beta with the market portfolio.

Interestingly, we can also derive CAPM "backwards" by assuming returns are driven by a single-index model:

$$R_i = \alpha_i + \beta R_M + e_i$$

and then thinking through what would happen in equilibrium. Just as we did in the previous chapter using the Traynor-Black method, we would argue that investors would form portfolios by diversifying away the residual risk as much as possible and choosing assets with positive alphas (and selling assets with negative alphas). The latter efforts to buy positive alpha stocks would drive up their prices and drive down their returns.

Once all of the alphas are competed away, investors will then relentlessly pursue diversification so that they all hold the market portfolio. And then we are back to CAPM, where the expected alphas are zero and investors hold the market portfolio.

CAPM Assumptions

At the start of this chapter we noted certain specific assumptions underlying the CAPM:

- All investors follow the methods outlined in previous chapters to select their optimal portfolio, attempting to optimize the trade-off between expected return and standard deviation.

- All investors make the same assumptions about the inputs to their portfolio optimization problem (they have homogeneous expectations regarding returns, variances and covariances).

- All assets are publicly traded so that investors can buy or sell risky assets in any amount and can borrow or lend at the risk free rate without limitation.

In fact, there are a few more assumptions that are needed for the CAPM:

- Investors' planning horizons are only one period.

The main reason we need to use this assumption is to rule out the possibility that investors would react to news about changes in *future* investment opportunities. If, for instance, investors cared about inflation, they might care about correlation with inflation and then wouldn't meet the core assumption we made that they care only about mean and standard deviation of returns.

- All information is publicly available. This helps to ensure that, in fact, investors are using the same input assumptions.
- There are no taxes. This ensures that the after-tax optimal risky portfolios are the same for everyone.
- There are no transaction costs, which might otherwise inhibit trading and complicate investors' attempts to hold their optimal risky portfolio.

Of course some of these assumptions are far from realistic. For instance, in reality short sales are somewhat restricted due to collateral requirements, a limited supply of shares and rules/regulations that prevent some investors from short sales. In the next section, we'll examine how violations of certain assumptions can impact the CAPM.

Extensions of CAPM

Relaxing some of the less realistic assumptions used for the standard CAPM can have interesting implications, which we'll explore below.

Zero Beta CAPM

The standard CAPM was derived by first assuming that investors can borrow or lend at a purely risk free rate, without restrictions. This meant that all investors ultimately owned portfolios that were a combination of the market portfolio and the risk free asset.

Suppose that unrestricted borrowing and lending were not possible. It turns out that much of what we did to derive CAPM could be modified slightly (see the Appendix immediately following the questions for this chapter) such that instead of selecting portfolios that are merely combinations of the risk free rate and the market portfolio investors select combinations of the market portfolio and a *zero beta portfolio*. The key result is that the expected return on any stock can be found to be similar to the CAPM but with the expected return on the zero beta portfolio replacing the risk free rate.

$$E(r_i) = E(r_z) + \beta[E(r_M) - E(r_z)]$$

Notice that the expected return on the zero beta portfolio will be higher than the expected return on the risk free asset. This will produce a SML with a higher intercept that could make it appear as though low beta stocks have positive alphas and high beta stocks have negative alphas.

Labor Income and Non-traded Assets

The standard CAPM assumes that the market portfolio contains all risky assets, which rules out the existence of future labor income (earnings from your job, referred to as human capital) or privately held businesses (non-traded assets).

Let's take non-traded assets first. If these have the same risks as traded assets, then the effect might not be that material but for some limited liquidity. But if the risks are different, then investors who own these non-traded assets may bid up the price of traded assets that act as a hedge against their risks, or similarly bid down the price of assets whose risks are highly correlated with the risks in these non-traded assets.

More problematic is how to handle labor income (human capital) because it is less portable and more difficult to hedge using traded securities. As a result, an individual's exposure to risks associated with their labor income could indeed impact their portfolio choices, such as their decision to avoid owning shares in their company or industry they work in, despite it being part of CAPM's efficient portfolio.

Using one possible adjustment to account for labor income, the *adjusted beta* would reflect not only the covariance of returns with the market but also the covariance of returns with the investor's human capital. We would expect this covariance to on average be positive, which would result in larger adjusted betas (and risk premiums) when the standard CAPM beta is lower than 1.0 and smaller adjusted betas (and risk premiums) when the standard CAPM beta is higher than 1.0. As a result the SML will be less steep. It will then appear as though there are positive alphas for low beta stocks and negative alphas for high beta stocks.

Multiperiod Model and the Intertemporal CAPM (ICAPM)

Another unrealistic assumption behind CAPM is that investors care only about risk and return in a single time period. It might be more realistic to consider a lifetime consumption plan and assume that investors will rebalance their portfolios as their wealth changes.

If we assume that risk preferences are constant, uncertainty about returns is the only source of risk and the return distributions do not change unpredictably over time, then the single period CAPM is still appropriate.

But consider the case where the underlying parameters such as the risk free rate, the market risk premium or the market risk can change over time. Since these changes could negatively impact an investor's level of future consumption, they will bid up the prices of assets that can be used to hedge against this risk. Put another way, rather than simply evaluate the degree to which an asset's return is correlated with the market portfolio, investors will also consider the degree to which an asset's return is correlated with changes in these variables. Assets whose returns will be higher when these variables change adversely will be more desirable, and hence have a higher price or lower expected return, than the standard CAPM predicts.

Another source of risk relates specifically to the real reason why investors want high returns — so they can consume goods and services! As a result, inflation risk will impact investor preferences and again they will bid up the prices of portfolios that can act as a hedge against inflation risk.

More generally, any number of risks other than just overall market risk may be appropriate to add to the standard CAPM, where we will be concerned with not just the beta relative to the market portfolio but also the betas for various hedge portfolios. The APT Model discussed in the next chapter is a broad generalization of this idea.

Consumption-Based CAPM

The standard CAPM assumes that investors simply care about increasing their wealth, but underlying that assumption is also the idea that the reason they are interested in this is so that they can consume goods and services. It stands to reason then that what might really matter to investors is whether an asset's returns are high or low when times are *good* or *bad* in terms of consumption (or, more formally, consumption growth). If my consumption is already high, I may not mind it too much if a risky asset's payoff is low; if my consumption is low though, then I might be even more sensitive to low payoffs. The important implication is that risk premiums will be driven by covariance with consumption as opposed to covariance with market returns.

Using this logic, it is possible to derive a model, referred to as the Consumption CAPM (CCAPM), which is quite similar to the standard CAPM except that the market portfolio is replaced by a portfolio that tracks consumption.

Liquidity

The issue of liquidity appears in other readings on the syllabus, but the context varies in each case and so some care is needed to understand how these concepts relate to each other. In this reading, the focus is on how investors take into account **liquidity costs** in the form of illiquidity discounts and **liquidity risk** arising from *changes* in the liquidity of an asset to determine the prices they are willing to pay for securities.

Liquidity Costs and Asset Prices

The *liquidity* of an asset is the ease and speed at which it can be sold at a fair market value in a timely fashion. It reflects the following three factors:

- the cost of engaging in the transaction, notably the *bid-ask spread*,
- the impact that your own selling will have on the price itself, and
- the immediacy or speed at which the transaction can occur.

An *illiquidity discount* can be thought of as the discount from fair market value that an investor would have to accept in order to sell an asset quickly and can be thought of as a cost of trading.

An important study by Amihud and Mendelson demonstrated the importance of liquidity for security markets. They found that returns for illiquid stocks are substantially higher than the returns for liquid stocks, but that large bid-ask spreads can make these stocks prohibitively expensive for investors who trade frequently. This suggests that if an investor intends to trade shares regularly, he should stick with liquid stocks with low bid-ask spreads. But if he does not intend to trade regularly, he might want to stick with illiquid shares, allowing him to capture the differential return without paying the bid-ask spread multiple times.

To begin to understand how bid-ask spreads can impact the prices of securities, note that market makers who buy and sell shares from investors have a lower price at which they are willing to buy shares (the bid price) than at which they are willing to sell shares (the ask price). The less liquid the stock (i.e. the fewer customer orders that come in regularly), the lower the bid price and the higher the bid-ask spread because of the risk the market maker faces while he holds the security in inventory. The size of the bid-ask spread represents a net cost to the investor, since they pay the ask-price to get the stock and then only receive the bid-price to sell it. The more often the investor trades, the more often he pays this cost and the lower his returns will be.

Another important factor has to do with information asymmetry. When I buy or sell an asset I have to be concerned with whether I am trading with someone who has more information than me and is “picking me off” by only accepting my bids or offers when I have mispriced the asset. The effect on the bid-ask spread depends on the mix of *noise traders* (liquidity traders) who do not have any specific private information (they may just be trading to rebalance their portfolio or to obtain funds for another purpose) and *information traders* who are trading based on private information. The more I am concerned that I am trading with an information trader who is capable of taking advantage of my lack of information, the more I will demand to be paid to sell and the less I will be willing to pay to buy — hence a larger bid-ask spread.

More importantly, since I know that on average I pay half the bid ask spread when I buy the security and half when I sell it, I will factor this cost into the price. But then again, so will the person I sell it to. As a result, if the stock is going to be sold once a year forever, the current price will actually reflect the present value of all of these bid-ask costs, which could amount to a significant reduction in the equilibrium price.

In the end, as noted above, investors who trade often will be more impacted by these trading costs than investors who trade infrequently and so each type of trader will likely self-select. Investors with shorter horizons will hold more liquid securities and investors with longer horizons will hold more illiquid securities. This clientele effect will dampen the effect the bid-ask spread will have on the illiquidity discount.

Liquidity Risk and Asset Prices

In addition, *changes in liquidity* introduce a new source of risk. Prices of securities whose liquidity is lower when it is most needed will require an additional price discount due to the systematic nature of the liquidity risk. One such model discussed in the text adds liquidity considerations to the standard CAPM in several ways. First, it includes the expected cost of liquidity as a component of the expected return, with an adjustment to reflect the amortization of this cost over the average holding period. Second, it measures the market risk premium net of the average market liquidity premium. And third, it adds three additional liquidity betas (with different signs) to reflect the following sources of systematic liquidity risk:

- *Sensitivity of the security's illiquidity to market illiquidity* — This reflects the fact that investors will want more compensation for liquidity risk if the security becomes illiquid when general liquidity is low.
- *Sensitivity of the stock's return to market illiquidity* — This reflects the fact that investors will accept a lower expected return on stocks whose returns are higher when market illiquidity is greater.
- *Sensitivity of the security illiquidity to the market's rate of return* — This reflects the fact that investors will expect a lower expected return on securities that can be sold more readily when the market declines and they are poorer (need more liquidity).

Is the CAPM Valid?

This is an interesting question that is surprisingly difficult to answer, for a few reasons:

1. CAPM assumes that the *market portfolio* contains all risky assets, but this hypothetical portfolio cannot actually be constructed. Since we use a broad index such as the S&P 500 as our proxy for the market portfolio, testing CAPM requires us to know whether this proxy portfolio is really mean-variance efficient.
2. CAPM is stated in terms of expected returns, but we can only observe actual returns. This means that we cannot confirm if the market portfolio's reward to variability ratio is the highest of any other portfolio.
3. Related to the last point, CAPM predicts a linear relationship between expected returns for a stock and the expected market risk premium. Since we cannot observe the expected returns, it is impossible to test the linear relationship.

The CAPM makes two closely related predictions — the market portfolio is efficient and the expected return-beta relationship depicted in the SML is accurate (i.e. the expected values of alpha are zero). However, because we cannot observe the true market portfolio that contains all risky assets, it is not possible to literally test the validity of CAPM.

We could, however, test the expected return-beta relationship (the SML). For instance, we could test using historical data whether returns are a linear function of betas, whether diversifiable risk affects returns and whether other sources of risk affect returns using a regression such as:

$$R_{i,t} = \lambda_0 + \lambda_1 \beta_i + \lambda_2 \sigma_{e_i}^2 + \nu_{i,t}$$

If the CAPM is correct, then we should find that:

- $\lambda_0 = 0$ (the alphas should be zero)
- $\lambda_1 = R_M$ (the slope of the SML should equal the market risk premium)
- $\lambda_2 = 0$ (diversifiable risk doesn't earn a risk premium)

So, how does the CAPM usually do in tests like this? As we will discuss in more detail in BKM Chapter 13, the early tests generally confirmed the linear relationship between expected return and beta, they confirmed that non-systematic risk does not affect expected return (i.e. beta is the only relevant variable), however the actual models were not fully verified because the intercepts of the tested models were too high and the slopes were too flat.

These latter two observations can be partly explained in terms of the Zero Beta version of the CAPM discussed above, but more recent work by Fama and French has identified additional factors besides the market beta that may be important to understand security returns.

Is this a problem that the CAPM fails these tests? First, all tests of CAPM suffer from parameter estimation error and so often these tests are inconclusive. For instance, Miller and Scholes did an analysis in which they generated samples of data by assuming that indeed the CAPM was valid and then performed tests of CAPM using this generated data. As in other studies, CAPM failed these tests even though the data literally assumed it was valid.

CAPM may fail these tests for other reasons as well. We often assume that the residuals are uncorrelated across firms, but indeed firms in the same industry may have correlated residuals. We also often assume that parameters are constant across time, but indeed they may vary over time or may vary in response to changing economic conditions.

CAPM and the Investment Industry

While academics have begun to abandon the CAPM in favor of models that can capture additional risk factors or better reflect time-varying parameters, practitioners still use CAPM quite widely. Part of why this is the case is that the empirical evidence has shown that it is very difficult for professional investors to beat a broad market index such as the S&P 500. If the S&P 500 can therefore be viewed as efficient and if alphas are essentially zero, then for many practical applications CAPM can be assumed to be valid.

Practice Questions

Question 1. Assume that you are going to combine an investment in the market portfolio with the risk free asset such that your risk, as measured as the standard deviation of returns, is equal to 5%. Use the capital market line to determine what your expected return would be for this portfolio. Assume that the expected return for the market is 15%, the risk free rate is 5% and the standard deviation of the market is 10%.

Solution. We know that when we combine the market portfolio with the risk free asset, our resulting returns lie along the capital market line, which goes through the risk free rate on the y-axis and the market portfolio point, (σ_M, r_M).

The equation for this line is easy to find, because it goes through the points $(0, r_f)$ and has a slope of $(r_M - r_f)/\sigma_M$. That tells us:

$$
\begin{aligned}
E(r_i) &= r_f + \frac{E(r_M) - r_f}{\sigma_M}\sigma_i \\
&= 5\% + \frac{15\% - 5\%}{10\%}(5\%) \\
&= 10\%
\end{aligned}
$$

Question 2. Using the same information from Question 1, confirm your result by first finding the proportions to invest in r_f and r_M such that the portfolio's standard deviation is 5% and then use those proportions to calculate the portfolio's expected return.

Solution. This is what we did earlier. Remember that the standard deviation for a portfolio is:

$$\sigma_p = [w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{12}]^{\frac{1}{2}}$$

If one of the assets is the market portfolio and one has no risk, then this simplifies to $\sigma_p = w_M\sigma_M$. So if we want a 5% portfolio standard deviation, then $w_m = 50\%$. Therefore, the expected return is $.5(5\%) + .5(15\%) = 10\%$, just as before.

Question 3. Assume that the risk free rate is 5% and the expected return for the market is 13%. If a stock has a beta of 2.0 and the CAPM applies, what is the expected return for the stock?

Solution. Using the information, the equation for CAPM is $E(r_i) = r_f + \beta[E(r_M) - r_f] = 5\% + 2(13\% - 5\%) = 21\%$.

Question 4. Assume that the risk free rate is 5% and the expected return for the market is 13%. If a stock has a standard deviation of 20%, has a correlation with the market of .8 and the standard deviation of the market is 40% what is the expected return for the stock? Assume that the CAPM applies.

Solution. This is just like the previous question, but now we have to find beta.

Remember that beta was defined originally as $\beta = \sigma_{iM}/\sigma_M^2$ and $\sigma_{iM} = \rho\sigma_i\sigma_M$. Therefore, $\sigma_{iM} = .8(.2)(.4) = .064$ and $\beta = .4$. Using the information, the equation for CAPM is $E(r_i) = r_f + \beta[E(r_M) - r_f] = 5\% + .4(13\% - 5\%) = 8.2\%$.

Question 5. Suppose the current price of IBM's stock is \$80, investors expect IBM's stock to equal \$88 in one year, IBM's beta is equal to 1.3, the risk free rate is 5%, the expected return on the market is 13%, and that CAPM applies. What do you think would happen to the current price of IBM's stock? What would cause this?

Solution. First, it is easy to see that according to the price information given the expected return for IBM is equal to:

$$\frac{E(P_1)}{P_0} - 1 = 10\%$$

According to the CAPM, the expected return should be:

$$E(r_i) = r_f + \beta[E(r_M) - r_f] = .05 + 1.3(.13 - .05) = 15.4\%$$

This means that investors who invest in assets with risk of $\beta = 1.3$ generally expect to earn 15.4% returns. In the case of IBM stock, they expect only 10%. Therefore, nobody would want to own IBM stock — it has an inferior return relative to its risk. If nobody wants to own IBM stock, clearly its price must fall. In fact, it will keep falling until people once again were willing to own it, which would occur when its expected return equaled 15.4%, or when its price dropped to 76.3.

Question 6. In Question 5, I said that IBM's expected return was inferior so nobody would want to own it. Show that investors could achieve a portfolio with a beta of 1.3 simply by investing some of their funds in the market portfolio and some in the risk free asset. Describe how they would do this and show that the expected return from this strategy matches the CAPM result of 15.4%.

Solution. Whenever you own a portfolio of stocks, the portfolio beta is just an average of the betas of its components, with the weights equal to the amount invested in each. Also, note that by definition, $\beta_M = 1$ and $\beta_{r_f} = 0$.

Therefore, if the portfolio beta is given as:

$$\beta_p = w_{r_f}\beta_{r_f} + w_M\beta_M = 1.3$$

then w_M must equal 130% and $w_{r_f} = -30\%$. Recall that this means we are going to borrow 30% (that's what it means to invest a negative amount in the risk free asset) of our initial portfolio value and invest that and the rest of our money in the market. The expected return would then be:

$$-30\%(5\%) + (130\%)(13\%) = 15.4\%$$

Question 7. List three reasons why it is hard to test the CAPM.

Solution.

1. you cannot observe the true market portfolio,
2. it is difficult to estimate the model parameters (expected return for the market portfolio, betas) accurately, and
3. the true model parameters are unlikely to remain constant over time.

Question 8. Tests of the CAPM often fail merely because the standard CAPM model is overly simplified. List and briefly describe four extensions of CAPM that at least partially explain the failure of these tests.

Solution. Four extensions discussed include:

i. Zero Beta CAPM — Rather than assume that investors select portfolios that are merely combinations of the risk free rate and the market portfolio, the zero beta version of CAPM assumes that investors select combinations of the market portfolio and a *zero beta portfolio*. The key result is then that the expected return on any stock can be found to be a similar to the CAPM but with the expected return on the zero beta portfolio replacing the risk free rate.

$$E(r_i) = E(r_z) + \beta[E(r_M) - E(r_z)]$$

This model results in a security market line that has a higher intercept and a lower slope, which is consistent with empirical tests that make it appear as though low beta stocks have positive alphas and high beta stocks have negative alphas.

ii. Labor Income and Non-traded Assets — The standard CAPM assumes that the market portfolio contains all risky assets whereas tests of CAPM are required to substitute a market index that is only a subset of this true market portfolio of all risky assets. Perhaps the CAPM is valid but the absence of assets such as investors' future labor income (earnings from their job) or privately held businesses (non-traded assets) create significant distortions? When these other sources of risk are taken into account, will tend to cause betas to be smaller than in the standard CAPM and as a result the SML to be less steep. This could explain the apparent positive alphas for low beta stocks and negative alphas for high beta stocks.

iii. Lifetime Consumption and the ICAPM — Another unrealistic assumption behind CAPM is that investors care only about risk and return in a single time period. It might be more realistic to consider a lifetime consumption plan and assume that investors care primarily with how asset returns impact this lifetime consumption. Rather than simply evaluate the degree to which an asset's return is correlated with the market portfolio, investors will also consider the degree to which an asset's return is correlated with changes in the risk free rate, the market risk premium, the market risk or inflation. Assets whose returns will be higher when these variables change adversely will be more desirable, and hence have a higher price or lower expected return, than the standard CAPM predicts.

iv. Consumption CAPM — Since investors really just care about consumption, a more appropriate source of risk premiums may be the covariance of an asset's payoffs with an index of consumption.

Question 9. BKM discuss an alternative to CAPM that explicitly takes into account the effect of liquidity costs and liquidity risk by measuring the beta with respect to the stock returns and the market returns *net* of their respective liquidity costs. Defining the stock's liquidity cost as

c and the market portfolio's liquidity cost as c_m and rewriting beta in terms of the net return covariance:

$$\beta = \frac{Cov(r - c, r_m - c_m)}{Var(r_m - c_m)}$$

Looking solely at the numerator for now, we can rewrite this as the sum of its four components:

$$Cov(r - c, r_m - c_m) = Cov(r, r_m) + Cov(c, c_m) - Cov(r, c_m) - Cov(c, r_m)$$

Describe what each of the four terms on the right-hand-side represent in relation to the betas described in the textbook.

Solution. Each of the four terms can be defined as follows:

$$\begin{aligned} Cov(r - c, r_m - c_m) = \overbrace{Cov(r, r_m)}^{\text{Market Beta}} + & \overbrace{Cov(c, c_m)}^{\text{Commonality in Liquidity}} \\ - & \overbrace{Cov(r, c_m)}^{\text{Return Sensitivity to Market Liquidity}} \\ - & \overbrace{Cov(c, r_m)}^{\text{Liquidity Sensitivity to Market Returns}} \end{aligned}$$

Notice that when these covariances are divided by the scale factor $Var(r_m - c_m)$ these become various betas, which include the traditional CAPM market beta and three specific liquidity betas:

- Commonality in Liquidity Beta — This reflects the premium required because a security may become illiquid when the market is illiquid. Empirically, most stocks' illiquidity is positively correlated with market illiquidity. Empirical studies (by Acharya and Pedersen) find that the return premium (product of the market risk premium and the beta) for this is small (0.08%) though.

- Return Sensitivity to Market Liquidity Beta — This reflects the fact that investors will prefer (will pay more and accept lower returns) securities with high returns when the market is illiquid. Empirical studies find that the return premium for this is 0.16%.

- Liquidity Sensitivity to Market Risk Beta — This reflects the fact that investors would pay more (be willing to earn a lower return) for a security whose liquidity was high when the market return is low, hence the negative sign in the above equation. Empirical studies find that this is the most important source of liquidity risk, with a measured return premium of 0.82%.

Appendix: Derivation of Marginal Risk

In the body of these notes, I suggested that the marginal risk for a portfolio invested some percentage, y, in the market portfolio and $1 - y$ in the risk-free asset would be as follows when a small amount, δ_i, of a particular stock is added to the portfolio:

$$\text{Contribution to Variance} = 2\delta_i y Cov(r_i, r_M)$$

To see how to derive this formula, we intend to borrow δ_i at the risk-free rate and invest it entirely in a particular stock. In that case, we would have y percent of our assets invested in the market, δ_i percent invested in Stock i and $1 - y - \delta_i$ percent invested in the risk-free asset.

Using the formulas for the variance of a combination of three assets but noting that the risk-free asset variance is zero and its covariance with each of the other two assets is also zero, then the variance of this new portfolio is:

$$\sigma_p^2 = y^2\sigma_M^2 + \delta_i^2\sigma_i^2 + 2\delta_i y Cov(r_i, r_M)$$

Now notice that when we originally only owned the market portfolio combined with risk-free borrowing or lending the risk was $y^2\sigma_M^2$. Therefore, the increase in risk is simply:

$$\delta_i^2\sigma_i^2 + 2\delta_i y Cov(r_i, r_M)$$

Since δ_i is small, then δ_i^2 is very small and we can ignore that term. Therefore, our risk increases by:

$$\text{Increase in risk for small } \delta_i = 2\delta_i y Cov(r_i, r_M)$$

Appendix: Zero Beta Version of the CAPM

This appendix shows the derivation of the Zero Beta version of the CAPM. The text doesn't derive this model in detail and so it is unlikely to be highly relevant for the exam. However, I believe that it will give you important insight into the Markowitz portfolio selection model and the overall derivation of CAPM. For this reason, I strongly encourage you to review the explanation below.

Review of the Basics

First, let's quickly review the standard CAPM. Start by assuming that we know the expected return and risk for every possible risky asset (referred to as stocks here). We know that if we create portfolios of risky assets, then there will be an infinite number of possible combinations of risk and return. As astute investors, we'll want to invest in only those that have the highest return for any level of risk and the lowest risk for any level of return. The combinations that meet this criterion will be a subset of all possible portfolios, called the efficient portfolios.

Now whether we choose a low risk, low return portfolio or a high risk, high return portfolio will depend entirely on our personal risk preferences.

But if we further assume other things, like everyone is free to borrow or lend at the risk free rate, that only risk and return matter to people, that everyone has the same information about risk and return, etc., then we can reach some important conclusions.

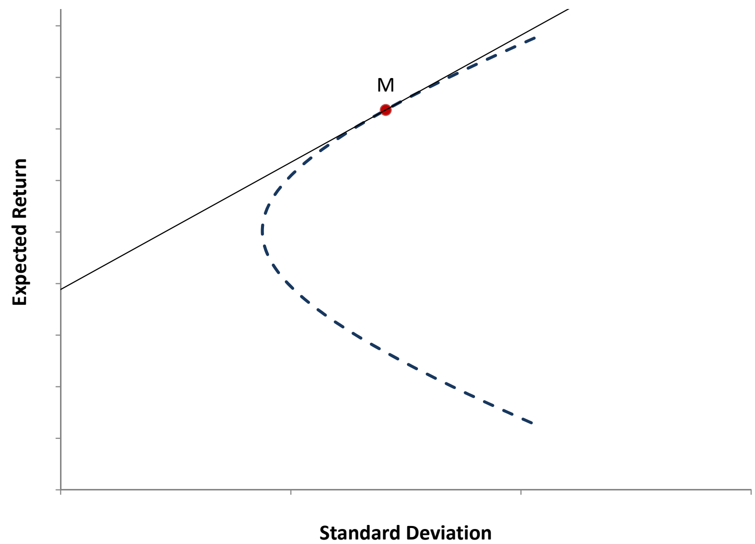
Figure 1 on the facing page shows the important result that when we let people borrow or lend, then combinations of a risky portfolio and the risk free asset will plot along a straight line. As a result, the only efficient portfolios are those that are a combination of the risk free asset and the risky portfolio which is labeled M in the graph.

As you can see, some people may want high risk, so they'll invest in M and will borrow at the risk free rate; others will prefer low risk and will invest in M and lend at the risk free rate. But the key thing is that everyone will want to own the same portfolio of stocks — Portfolio M.

And if they all want the same portfolio, then that portfolio must be the market portfolio. Further, note that there are only 2 assets that people own. They either own the market portfolio or the own the risk free asset. That doesn't mean that they can't own a bit of both, but their portfolio is only a combination of these two assets/portfolios

So what does that tell us about the returns for any individual stock? Well, not much, yet. So far we've been talking about expected return and the standard deviations of returns. What if we were to define the term Beta and use that as our risk measure (let's just skip over why this might make sense)?

FIGURE 1. Efficient Frontier with Risk Free Asset



By definition, the beta for the risk free asset is zero and the beta for the market portfolio is 1. Remember that we just concluded that the only portfolios we might want to own is either:

a. the market portfolio,

b. the risk free asset, or

c. some combination of the two.

If we own the market portfolio, our return would be r_M and our beta would be 1. If we owned the risk free asset, the return would be r_f and the beta would be 0. And if we had some combination of the risk free asset and the market portfolio, then the beta for our portfolio would be an average of 0 and 1, based on the proportions we had in each and the expected return would be an average of the two expected returns.

If we were to plot all of these returns and betas for various combination of the risk free asset and the market portfolio, then the risk and return would all plot along a straight line:

$$r_i = r_f + \beta(r_M - r_f)$$

This is the standard CAPM.

Note the subtle point. We began discussing a graph with expected return and standard deviation, used that to conclude that the only two things people would own is either the market portfolio or the risk free asset and then came up with a formula relating expected return and beta.

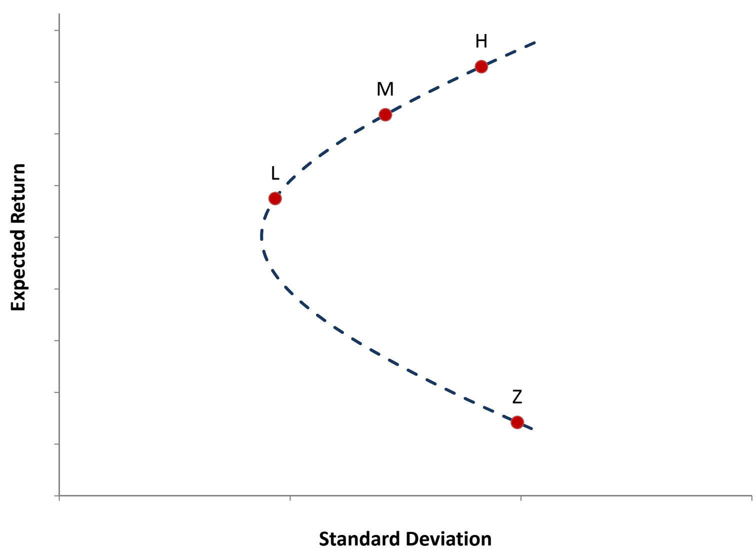
But that relationship is just the relationship for portfolios that are efficient — what does it say about individual stocks? Well, think about what would happen if IBM had a beta of 1.3 and expected returns that were lower than the CAPM line suggests. Since people are free to invest in the risk free asset and the market to get a beta of 1.3 with higher expected returns, clearly nobody would want to own IBM stock. And if nobody wanted to own it, the price would have to fall and the expected return would naturally rise. That's why the CAPM line applies to all stocks.

Zero Beta Version of CAPM

Now think about an extreme case where people are not allowed to invest in the risk free asset. That is, they cannot borrow or lend at the risk free rate. Let's follow the same thought process in that case.

It starts the same — people plot all combinations of stocks and pick the portfolio that best suits their risk preference. Without borrowing and lending, we no longer conclude that they all want to own the same portfolio, but they all are still smart enough to limit their choices to efficient portfolios. So some people will pick a low risk portfolio, *L*, and some a high risk portfolio, *H*, both shown in Figure 2. Since some people own *L* and some people own *H*, then

FIGURE 2. Efficient Frontier with No Risk-Free Asset



we know that the aggregate market portfolio, which contains all of *L* and all of *H*, must lie somewhere on the graph. But where?

It turns out that any combination of two minimum variance portfolios (any portfolio on the upper or lower portion of the curve above), like *L* and *H* in my graph, is also on the minimum variance curve. Therefore, we know that *M* lies between *L* and *H* and is on the efficient frontier (we'll get back to this in a moment).

So now we have a world in which investors either own L or H. And they limit their choices to these two portfolios because these are the best portfolios that are efficient and are consistent with their risk preferences. Note though that I made it simple by saying that these are the only two portfolios because people only choose high risk or low risk portfolios. If we also assume that other investors like medium risk, or medium high risk, etc. then they'll perhaps own different efficient portfolios. But in all cases, they'll be somewhere on the efficient frontier.

Again without proving it, I will state that there exists some portfolio, Z, that is on the lower portion of the minimum variance frontier and that is uncorrelated with portfolio M. This is based on a point that is only briefly mentioned in the textbook that *every* portfolio on the efficient frontier has a companion portfolio that is on the lower portion of the minimum variance frontier and that is uncorrelated with it. We are just carefully selected portfolios M and Z here.

Further, for the same reason that we knew we could construct M as a combination of L and H, we also know that we can construct portfolios L or H by combining the M portfolio with the Z portfolio.

Why do we need to know this? Because it means that investors don't have to choose between L and H, they can simply chose between M and Z — since L and H can be constructed from M and Z.

Just as we had before, once we know that investors limit their choices to just two portfolios—in the standard CAPM it is the risk free asset and M and now it is Z and M—then we can move away from the graph with expected return and standard deviation and think about the graph of expected return and beta. In this case, we know that some investors may choose portfolio Z with expected return of r_Z and beta of 0 (recall that I said Z and M were uncorrelated, so that means that they have zero covariance and therefore Z has zero beta). Other investors choose M, with expected return of r_M and beta of 1.0 (by definition). Any combination of M and Z that investors choose will therefore have a return that is an average of the two returns and a beta that is an average of the two betas.

Here, if the weight in the market portfolio is given as w then the expected return is:

$$r_i = wr_M + (1 - w)r_Z$$

Similarly, the beta is:

$$\beta = w\beta_M + (1 - w)\beta_Z$$

Since $\beta_M = 1$ and $\beta_Z = 0$ by definition, then $w = \beta$.

Plugging this into $r_i = wr_M + (1 - w)r_Z$ and rearranging, we have the zero-beta version of CAPM:

$$r_i = r_Z + \beta[r_M - r_Z]$$

As before, this applies only to portfolios that people choose, but if we look at any particular stock, then we can see that if it did not also lie on the same line, then investors would consider it to be inferior to their other options of some weighting of M and Z and therefore its price and return will adjust until its risk and return combination did lie along the line.

Since the equation for this line looks just like CAPM except that instead of the risk free asset we use an asset that is uncorrelated with the market and thus has a zero beta, we call this the Zero Beta version of the CAPM.

Please note the following additional points:

- The "M" portfolio I mentioned above for the zero beta model above is also the Market Portfolio, since it was stated that it represented the aggregate of the L and H portfolios.

- Note that nobody would actually own the Z portfolio on its own, since it is itself inefficient. However, people who want to own L will own some combination of M and Z. Similarly, people who want to own H would own some combination of M and Z, but in this case they'd own a negative amount of Z (they would sell Z short). The aggregate net holdings of Z would be zero.

BKM Chapter 10: APT and Multi-Factor Models

This chapter presents the Arbitrage Pricing Theory (APT), which looks similar to the standard CAPM. However, the logic and assumptions used to derive this model are completely different from those behind the CAPM.

Factor Models

Recall in Chapter 8 we discussed single factor models that assumed the actual returns for an individual security consisted of some amount of expected return, an additional return (with expected value of zero) based upon an economy-wide shock denoted by F (note that F represents a shock and so it is measured as the deviation of a variable from its expected value) and a firm-specific random component. We wrote that factor model equation as:

$$r_i = E(r_i) + \beta_i F + e_i$$

We also saw that we could write this formula in terms of *excess* returns, denoting excess returns with R_i instead of r_i:

$$R_i = E(R_i) + \beta_i F + e_i$$

This can easily be extended to include multiple factors in an obvious fashion. For instance, suppose that the two most important macroeconomic sources of risk are unanticipated changes in GDP (the business cycle) and unanticipated changes in interest rates. We might write this as:

$$R_i = E(R_i) + \beta_{i,\text{GDP}}\text{GDP} + \beta_{i,\text{IR}}\text{IR} + e_i$$

The betas in the equation above, referred to as *factor loadings* or *factor betas* when used in a multifactor model, reflect the degree of sensitivity of the excess return to the factors.

What Do Factor Models Tell Us About Expected Returns?

Students are often tempted to take the expected value of the factor model equation shown above in order to say something useful about the expected returns. But note that the expected return already appears on the right-hand side of this equation!

Instead, what we will do below is simply assume that this factor model is correct and then, along with other assumptions, see whether this can lead us to a useful equation for the expected (excess) returns on risky assets.

Multifactor SML

Recall that the Security Market Line (SML) is an equation for the the expected return, of the following form:

$$E(r_i) = r_f + \beta_i[E(r_M) - r_f]$$

which can also be written in terms of the excess return and the market *risk premium*

$$E(R_i) = \beta_i \mathrm{RP}_M$$

For obvious reasons, it would be relatively trivial to extend this to include multiple sources of risk premiums that result from sensitivity to multiple risk factors.

An important point to keep in mind is that the multifactor *factor model* and the multifactor SML are NOT the same thing. The former is a statement about what causes actual returns to deviate from their expected values; the latter is a statement, based on some underlying theory, of what the expected returns are. CAPM provided us with one theory for the expected returns and risk premiums. The APT will provide us with another theory.

The Fundamental APT Formula

The APT equation for the expected return on a risky portfolio of assets looks almost identical to a multifactor CAPM:

$$E(R_i) = \beta_1 \mathrm{RP}_1 + \beta_2 \mathrm{RP}_2 + \cdots + \beta_N \mathrm{RP}_N$$

The only differences between that formula and the CAPM formula for the expected return is that there may be more than one risk factor and none of the risk factors need to be the market index return. We will see shortly that there are important differences in the logic used to derive the formulas, but let's hold off on that issue for now.

As an example of how the APT would be used, suppose I determined that there are only five risk factors (i.e. five sources of systematic risk). The table below summarizes the average risk premium earned (as % per year) for each of these five risk factors as well as the sensitivity of a particular portfolio to each of the factors.

TABLE 1. APT Betas and Factor Risk Premiums

Risk Factor	Beta	Factor Risk Premium (%)
Confidence	0.27	2.59
Time Horizon	0.56	-0.66
Inflation	-0.37	-4.32
Business Cycle	1.71	1.49
Market Timing	1.00	3.61

Given this information, we can determine the expected excess return for this portfolio using the APT formula shown above:

$$\begin{aligned} E(R_i) &= \beta_1 \mathrm{RP}_1 + \beta_2 \mathrm{RP}_2 + \cdots + \beta_N \mathrm{RP}_N \\ &= .27(2.59\%) + .56(-.66\%) - .37(-4.32\%) + 1.71(1.49\%) + 1.00(3.61\%) \\ &= 8.09\% \end{aligned}$$

That should have been trivial and should have seemed like a multifactor CAPM. The key to the APT is in the logic that is used to derive the formula we just used, starting from just two fundamental assumptions:

- Assumption 1 — Returns are generated by a factor model. This means that actual returns for any portfolio consist of its expected return, plus some return due to sensitivity to unexpected shocks that are common across all portfolios, plus an idiosyncratic risk term that is uncorrelated with the shocks.

- Assumption 2 — All assets can be bought or sold in any quantities at the market price, so in equilibrium arbitrage profits cannot be earned (see below).

What follows now is the derivation of the formula we used. Since the final formula is so simple to use, to answer anything other than the most trivial exam questions you need to understand this derivation.

Arbitrage

The concept of arbitrage is critically important in finance. An arbitrage opportunity exists when a zero investment opportunity (one that requires no upfront cash outlay) produces a sure profit (risk-free). Because there is no risk, we would expect those kinds of opportunities to be exploited by investors regardless of their risk preferences. These profits cannot last long because investors will take steps to exploit them — forcing up the prices of some assets and forcing down the prices of others.

This is very different from the dominance arguments we used earlier, which suggested that one asset might be superior to another but only subject to a particular level of risk aversion.

Short Selling

A critical element of exploiting arbitrage opportunities is the ability to buy or sell any asset, in any quantity, at the market price. This includes being able to sell an asset that you do not own. This is known as *short selling*.

Selling an asset *short* is accomplished by first borrowing the asset from someone, selling it immediately later buying it (from someone else) so that it can be returned to the person you originally borrowed it from. For instance, if you thought the price of an asset was going to drop, you might want to short sell the asset now to collect the (high) current price and then buy it later at the lower price before returning it to the person you initially borrowed it from. This “sell high, buy low” strategy would of course be profitable if you were right and prices did fall before you had to return the stock.

Nonetheless, short selling need not always be used solely to profit when you think prices will fall. It could merely be used to create a negative exposure to price increases for a particular asset. This would be done to hedge risks or to take advantage of arbitrage opportunities. As an example, if you thought that two assets should have the same price but for some reason do not, you may want to buy the cheaper of the two and short sell the more expensive of the two. This will give you a positive up-front payment and then eventually when the prices converge to be the same you can sell the one you own and use the money to buy the one you sold short (so it can be returned to the person you borrowed it from). Nothing in this strategy requires that the price of the asset sold short has to fall in an absolute sense.

In what follows, I will always assume that you can buy or sell any asset at its market price (in any quantities).

Law of One Price

The concept that an arbitrage opportunity cannot last long is often thought of in terms of the Law of One Price. Under the Law of One Price, two assets that are identical in all economically relevant respects should have the same price. If not, then we would say that an arbitrage opportunity exists and investors would be able to simultaneously buy and sell *identical* assets at two different prices. By buying at the low price and selling at the high price, investors could earn a net profit equal to the difference as a risk free profit. This activity would put upward pressure on the low-priced version and downward pressure on the high-priced version, until their prices were indeed equal.

One way to think of the Law of One Price is that the price of a Happy Meal has to be the same as the price of the burger, fries and drink. Of course, that isn't an ideal example because to exploit an arbitrage opportunity you need to be able to sell things you don't own. It would be hard to exploit this opportunity because you cannot walk into a McDonald's, buy the Happy Meal and simultaneously sell the burger, fries and drink separately to other customers. But if that were allowed, this is how an arbitrage opportunity would be exploited. This sort of activity would force the prices of the Happy Meal and the burger, fries and drink combination to have the same prices.

Other examples of the Law of One Price include instances where the same asset is available at different prices but in different locations. So long as the transaction costs were negligible, we wouldn't expect the price of a share of stock to be different if bought at the New York Stock Exchange as opposed to on the NASDAQ electronic exchange.

Other Forms of Arbitrage Opportunities

The Law of One Price relates to economically identical assets. But the concept of arbitrage opportunities also applies, for instance, if two assets have different prices but identical risks. By buying the one that sells at a low price and shorting the one that sells at a high price, we create an up-front profit with no net risk — the risks completely offset each other. In this case, we don't require the *expected* cash flows of the two assets to be identical, only that the risks associated with the cash flows be identical. For instance, the factor loadings on the two assets are identical even if the expected returns are not. We will actually use this version of an arbitrage opportunity below.

APT and Well Diversified Portfolios

We begin with a simple version of Arbitrage Pricing Theory (APT) assuming that returns are generated from a single factor model.

Outline of Key Ideas

Assume a single factor model can be used to explain the realized excess returns for securities such that:

$$R_i = E(R_i) + \beta_i F + e_i$$

Note that for a large portfolio with N stocks, each with weights $1/N$ in our portfolio, the firm specific risk is diversified away and therefore only the systematic risk should matter to us. For convenience, we expand our definition of *well diversified* to include portfolios that may not be equally weighted but are still large enough so that the unique risk is essentially zero and can be ignored. As a result, the portfolio returns can be written as:

$$R_p = E(R_p) + \beta_i F$$

Now notice what would happen if we had two well diversified portfolios, A and B, with different expected excess returns but the same betas (that is, the same degree of systematic risk). For example, assume that $E(R_A) = 10\%$ and $E(R_B) = 9\%$ but both had equal betas in the single factor model of $\beta_A = \beta_B = 1.0$.

Intuitively we can see that this probably couldn't last for long. Since well diversified portfolios are free of any non-factor risk, these two portfolios have equal risk but unequal returns. Whatever the realized value for F turns out to be, Portfolio A would always have a higher actual return than Portfolio B. This would entice investors to want to own A and not B, pushing up the price of A and pushing down its expected return (and the reverse for B).

Consider this again but with respect to a transaction designed to exploit this *arbitrage opportunity*. Suppose we invested \$1 million in A and sold (short) \$1 million of B. The actual returns in dollars for each of these portfolios and for the aggregate portfolio would be as shown in Table 2.

TABLE 2. Example of Arbitrage Opportunity

Portfolio A:	$[.10 + 1(F)](\$1{,}000{,}000)$
Portfolio B:	$-[.09 + 1(F)](\$1{,}000{,}000)$
Net Portfolio:	$0.01(\$1{,}000{,}000)$

Since this gives us a positive return of \$10,000 with no up-front net investment and no net risk (the weighted average beta is zero), arbitrageurs would try to exploit this by buying Portfolio A and selling Portfolio B simultaneously, pushing down the returns for Portfolio A and pushing up the returns for Portfolio B.

Risk Premiums

Another way to state this result is to say that if we combined two portfolios whose returns followed a single factor model in the proportions that eliminate all of the factor risk (produce an overall beta equal to zero), then the combination is risk free and should earn the risk free rate of return.

In the example above, the portfolios have the same betas, so it was obvious how to combine them to produce a zero beta portfolio. But more generally, if the betas are not equal they can still be combined to produce a zero beta portfolio using the following weights on Portfolios A and B:

$$w_A = \frac{\beta_B}{\beta_B - \beta_A} \qquad w_B = -\frac{\beta_A}{\beta_B - \beta_A}$$

The combined weights add to 1.0 and the combined excess return would be:

$$E(R) = \frac{\beta_B}{\beta_B - \beta_A}E(R_A) + \frac{-\beta_A}{\beta_B - \beta_A}E(R_B)$$

Setting this equal to zero (since the weighted average beta is zero and hence there's no systematic risk), we can rearrange terms to find:

$$\frac{E(R_A)}{\beta_A} = \frac{E(R_B)}{\beta_B}$$

This is the crucial result! It shows that since we would be able to combine two well diversified portfolios together and take advantage of arbitrage opportunities, **in equilibrium all well diversified portfolios must have the same risk premiums relative to their betas.**

Relationship to CAPM

From the previous point about the relative risk premiums for diversified portfolios, consider a reference portfolio such as the market portfolio and assume that this market portfolio represents the systematic risk factor in a single factor model. In this case, $\beta_M = 1$ by definition and therefore the previous relationship (which showed that excess returns relative to betas have to be equal across all well-diversified portfolios, can be written with the market portfolio parameters and then rearranged:

$$\frac{E(R_A)}{\beta_A} = \frac{E(R_M)}{1}$$

$$\Rightarrow E(R_A) = \beta_A[E(R_M)]$$

$$\Rightarrow E(r_A) = r_f + \beta_A[E(r_M) - r_f]$$

This is the same result as we developed for CAPM, but we did not need to make the same restrictive and unrealistic assumptions. We merely assumed that a single factor model determined security returns, that arbitrage opportunities would be exploited and that the portfolios in question were well-diversified so that we could ignore the residual risk.

APT vs. CAPM for Individual Assets

One weakness of what we just did is that it applies only to well diversified portfolios. What about less diversified portfolios (like most actively managed equity portfolios) or even individual assets?

We cannot use the same arguments to ensure that the residual risk is insignificant, however we can take some comfort in the fact that our APT estimate of the expected return may still be a useful and unbiased estimate of the expected return for most assets and portfolios. This isn't perfect, and it may be problematic that we can't rely on arbitrage arguments alone to ensure that portfolios or assets with high degrees of residual risk follow the APT formula. But it is nice that we can obtain a model of expected return without requiring the kinds of assumptions needed for the CAPM, such as the assumption that all investors use mean-variance analysis to optimize the Sharpe ratio of their portfolios.

APT also has an advantage over CAPM in that the APT is anchored by observable portfolios, while CAPM is based on the unobservable portfolio of all risky assets. And, as we will explore further in BKM Chapter 13, when we use a broad market index for the CAPM in place of the theoretical portfolio of all risky assets, we can't really be sure the CAPM is an unbiased predictor of expected return for individual assets anyway.

The APT and Portfolio Optimization in a Single Index Market

Recall in Chapter 8 we used the Treynor-Black procedure to find an optimal portfolio, which amounted to maximizing the information ratio, $IR = {}^{\alpha_A}/_{\sigma(e_A)}$, of the active portfolio. The information ratio for the portfolio is maximized when the weight on an individual asset is proportional to its information ratio.

Under the APT, we assume that for well diversified portfolios the residual risk is zero and that whenever you spot an arbitrage opportunity you would scale up your holding in that asset without bound, thus creating the pressure on prices that results in the equilibrium the APT predicts. The Treynor-Black procedure would result in the same, infinite, allocation to an asset with positive alpha and no residual risk.

The difference between the two is that the APT doesn't have anything to say about the case where the residual risk is not zero. However, the Treynor-Black procedure would dictate the same aggressive tilting of the portfolio towards positive alpha assets, but with a recognition of the limits of diversification. This puts the same price pressure on assets whose returns suggest a positive alpha, pushing returns towards those predicted by the APT.

Multifactor APT

While the APT does not specify how many factors there may be or even what they might be, it is easy to imagine a number of factors which could affect stock returns, like interest rates, inflation, oil prices, etc. Extending our single factor APT arguments to the case where many factors impact returns is rather trivial.

Here, we assume that there is a multifactor model where each factor represents the unexpected surprise in N systematic variables:

$$R_i = E(R_i) + \beta_1 F_1 + \beta_2 F_2 + \cdots + \beta_N F_N + e_i$$

We then proceed in the same fashion as for a single factor model, which results in a multifactor version of the APT formula for the expected excess return on a well diversified portfolio:

$$E(R_i) = \beta_1 \mathrm{RP}_1 + \beta_2 \mathrm{RP}_2 + \cdots + \beta_N \mathrm{RP}_N$$

Factor Portfolios

When working with multifactor APT models, there's an important simplification that will help us. Since we have so many assets to choose from, think about constructing ***factor portfolios*** that are well diversified portfolios that have factor sensitivities (betas), of 1.0 to only one factor and factor sensitivities of zero to all other factors.

For instance, in a two factor model, Factor Portfolio 1 has betas of $\beta_1 = 1.0$ and $\beta_2 = 0.0$ and Factor Portfolio 2 has betas of $\beta_1 = 0.0$ and $\beta_2 = 1.0$. In addition, we can define the return on a portfolio with no systematic risk, hence $\beta_1 = 0.0$ and $\beta_2 = 0.0$, as the risk-free rate.

Given the expected returns for these two factor portfolios and the risk free rate, we can easily determine the risk premiums for each of the factor portfolios. Then, given the factor sensitivities of any particular portfolio, we can easily determine that the total risk premium for this portfolio must equal the sum of its factor sensitivities times the associated factor portfolio risk premiums.

Numerical Example

Let's stop and look at a numerical example. Assume the risk free rate is 5%, the expected returns for the factor portfolios are $E(r_1) = 13\%$ and $E(r_2) = 9\%$ and you are interested in the expected return for a well diversified portfolio, Portfolio D, with factor sensitivities of $\beta_1 = 0.75$ and $\beta_2 = 1.5$.

In this case, the factor risk premiums are $13\% - 5\% = 8\%$ for factor 1 and $9\% - 5\% = 4\%$ for factor 2. Therefore, the total risk premium for this portfolio should equal the sum of the factor betas multiplied by the respective factor risk premiums and the expected return should equal the sum of the risk premium and the risk free rate:

$$\begin{aligned}
\text{Risk Premium} &= .75(8\%) + 1.5(4\%) \\
&= 12\% \\
\text{Expected Return} = E(r_D) &= r_f + \text{Risk Premium} \\
&= 5\% + 12\% \\
&= 17\%
\end{aligned}$$

Why does this have to be the case? It's because of the *arbitrage* that would be created if it were not the case. Assume that in fact the expected return for this portfolio were only 16%. Then consider the following arbitrage:

- Invest \$750,000 in Factor Portfolio 1,
- Invest \$1.5 million in Factor Portfolio 2,
- Borrow \$1.25 million at the risk free rate (this is the same as investing \$-1.25 million in the risk free portfolio).
- Sell short \$1 million of Portfolio *D*

The first three components represent a combination of the three key factor portfolios (treating the risk free asset as a particular factor portfolio with no factor risk). The net investment for this portion of the strategy will be \$1 million, with weights of .75 on Factor Portfolio 1, 1.5 on Factor Portfolio 2 and -1.25 on the risk free asset. This portfolio will have an expected return of $.75(13\%) + 1.5(9\%) - 1.25(5\%) = 17\%$ and will have average betas of $\beta_{P1} = .75$ and $\beta_{P2} = 1.5$.

Because we also sold short \$1 million of Portfolio *D*, we have zero net investment and, as shown below, zero factor sensitivities as well:

$$\begin{aligned}
\text{Combination of Factor Portfolios} &= \$1\text{M}[E(r_P) + \beta_{P1}F_1 + \beta_{P2}F_2] \\
&= \$1\text{M}[17\% + .75F_1 + 1.5F_2] \\
\text{Portfolio D} &= -\$1\text{M}[E(r_D) + \beta_{D1}F_1 + \beta_{D2}F_2] \\
&= -\$1\text{M}[16\% + .75F_1 + 1.5F_2] \\
\text{Net Portfolio Payoffs} &= \$1\text{M}[1\%] \\
&= \$10{,}000
\end{aligned}$$

Notice that this return does not depend upon the realization of the factors. Therefore, it represents a risk free profit that arbitrageurs would try to exploit.

This would increase the cost of purchasing the combination of factor portfolios and lower the cost of Portfolio *D* (because of all the efforts to sell it). The effect of this will be to lower the returns on the factor portfolios and increase the returns on Portfolio *D* until the arbitrage opportunity disappears.

APT's Formula for the Expected Return

Notice something subtle in the numerical example we just walked through. I began by assuming a two factor model determined excess returns:

$$R_D = E(R_D) + \beta_1 F_1 + \beta_2 F_2 + e_D$$

I was then able to determine, using arbitrage arguments, the expected return for Portfolio D without having any information about the values for F_1 or F_2. In other words, I never actually used the formula that I assumed determined returns; I never actually used the factor model. Instead, I set up an arbitrage argument that told me that once I knew the returns for the risk free asset and the factor portfolios, I could determine the expected return on any portfolio once I knew its factor sensitivities.

The portfolio I constructed to match the factor sensitivities of Portfolio D told me exactly what the return had to be and the arbitrage argument proved this had to be true. But let's rewrite that expected return formula for the combination portfolio I constructed and work backwards to rewrite it entirely using the variables associated with Portfolio D and the factor risk premiums, RP_i:

$$\begin{aligned}
E(r_D) &= \text{Expected Return on Combination of Factor Portfolios} \\
&= .75(13\%) + 1.5(9\%) - 1.25(5\%) \\
&= \beta_{D1}E(r_1) + \beta_{D2}E(r_2) + (1 - \beta_{D1} - \beta_{D2})r_f \\
&= r_f + \beta_{D1}[E(r_1) - r_f] + \beta_{D2}[E(r_2) - r_f] \\
&= r_f + \beta_{D1}\text{RP}_1 + \beta_{D2}\text{RP}_2
\end{aligned}$$

Notice that this final form simply says that the expected return on a well diversified portfolio is equal to the risk free rate plus its sensitivity to Factor 1 times the Factor 1 Risk Premium plus its sensitivity to Factor 2 times the Factor 2 Risk Premium, where the Factor i Risk Premium is the expected excess return for the Factor i Portfolio with $\beta_i = 1.0$ and $\beta_j = 0.0$.

Fama & French 3-Factor Model

The APT is elegant, but gives no guidance as to what the factors are or even how many there are. Since we do not need to know the expected values of the factors, in some ways we never need to know what they are to use APT. However, whatever the factors actually are, they should reflect systematic risks that are reasonably likely to lead to risk premiums — factors that concern investors enough to demand a risk premium to bear exposure to those sources of risk.

The Fama-French Three Factor Model is the most important, and most widely used, APT-style model. It was developed to help understand why small stocks (measured in terms of the total market value of the firm's equity) tended to earn higher returns than large stocks and why firms with high ratios of book value to market value also tend to earn higher returns than firms with low book to market ratios.

Rather than identify the specific risk factors that might be driving these empirical results, they created two portfolios, SMB and HML, that they argued could serve as proxies for unidentified risk factors (along with the market return as a third factor):

- SMB is a portfolio whose returns reflect the difference in returns for Small stocks and Big stocks

- HML is a portfolio whose returns reflect the the difference in returns for stocks with High ratios of book value to market value and Low ratios of book value to market value

They used a three-factor model, with the excess market return, SMB and HML as the three factors and estimated sensitivities of returns for different portfolios to these factors. Using the same logic as we used earlier for the APT model, they then were able to develop an expected return equation with these factor loadings and the expected risk premiums for these three factors:

$$E(r_i) - r_f = b_i[E(r_M) - r_f] + s_i E(\text{SMB}) + h_i E(\text{HML})$$

The FF model is not technically assuming that the firm size and book to market ratio are actually the specific risk factors that result in higher returns, as this would seem somewhat inconsistent with investor rationality. More subtly, they are assuming that the SMB and HML time series serve as proxies for some other unidentified risk factors. High book-to-market firms are likely to be in financial distress and small firms are more sensitive to business conditions, so it seems plausible that these indices are capturing risk factors that exist in the macroeconomy.

The danger is that these variables may fit the historical data by chance, especially given how easy it is for researchers to run regressions on all sorts of data until they stumbled on a model that fit the historical data. This was Black's initial critique of the FF work. But the FF results have held up in subsequent tests and there is evidence that the SMB and HML indices are indeed capturing meaningful macroeconomic effects. For instance, they have been found to be useful in predicting GDP growth in different countries.

We will return to the Fama-French model in detail in BKM Chapter 13.

Applying the APT

Suppose you wanted to actually use a model like the APT to estimate expected returns. How would you proceed?

The Two-Pass Regression Approach

In BKM Chapter 13, we'll discuss a two-pass regression procedure used by researchers to test the CAPM and the APT. But I think it makes sense to present it here in the context of the APT.

Recall that under the APT, we argued that if a multifactor model (with N factors) drove returns on well-diversified portfolios then the expected returns (on well diversified portfolios) would follow:

$$E(r_i) = r_f + \beta_{i1}[E(r_1) - r_f] + \beta_{i2}[E(r_2) - r_f] + \cdots + \beta_{iN}[E(r_N) - r_f]$$

Let's outline how you would parameterize this equation using the approach outlined in the Box titled "Using the APT to Find the Cost of Capital" in the BKM textbook.

- Step 1 - Identify plausible factors, such as unanticipated changes in interest rates, inflation, GDP, etc. and construct time series of these factors.
- Step 2 - Using a sample of firms (or portfolios), regress their historical returns against these indices to estimate the betas or factor loadings. This is called a first-pass regression because its goal is to just measure the sensitivities and will be performed once for each stock or portfolio.
- Step 3 - Now, as a second-pass, regress the returns for each stock on the estimated betas to get estimates of the risk premiums for each factor.
- Step 4 - Finally, do a similar first-pass regression to get the betas for the portfolio in question and then use these betas and the Step 3 risk premiums to parameterize the APT.

The Linear Algebra Approach

Although not discussed in the BKM textbook, there's sort of another way to use the APT, which I will refer to as the Linear Algebra approach because it involves simply solving a linear algebra problem. I am presenting it because this how the APT was tested on some of the older CAS exam problems and if you come across those questions it's helpful to understand them.

Remember that the formula for the expected excess return under the APT was obtained simply by assuming that a multifactor model drove actual returns and that arbitrage opportunities did not exist in equilibrium. Suppose you are told that a two-factor model drives returns. Under these assumptions, expected excess returns for all well-diversified portfolios are then just:

$$E(R_i) = \beta_{i1}[\text{RP}_1] + \beta_{i2}[\text{RP}_2]$$

Now suppose I told you the expected excess return for any **two** well-diversified portfolios, with known betas, and asked you to find the expected excess return for any other portfolio. This is a trivial linear algebra problem because the two portfolios given could be used to solve two equations with two unknowns (the unknown risk premiums, RP_1 and RP_2). Then, those risk premiums could be used along with the betas given for any other portfolio to calculate the expected return.

If instead of the excess returns you were given the total returns and you didn't already know the risk-free rate, then the algebra is a bit harder. In this case, you would need to know the returns and betas for three well-diversified portfolios and then solve three equations for the three unknowns. But it is still just a linear algebra problem.

And what if you are told that there are 3, 4 or more factors and not just 2? You can always solve for all risk premiums so long as you have the expected excess returns and betas for the same number of portfolios as you have factors (or one more if you need to know the risk free rate too).

Practice Questions

Question 1. Suppose returns on risky portfolios are driven by a two-factor model with unanticipated growth in GDP and changes in interest rates are the macroeconomic sources of risk for which risk premiums are generated:

$$R_i = E(R_i) + \beta_{i\mathrm{GDP}}\mathrm{GDP} + \beta_{i\mathrm{IR}}\mathrm{IR} + e_i$$

For a particular portfolio that you are interested in, you estimate the parameters of this factor model as follows:

$$R = 9\% + 1.4\mathrm{GDP} - 0.3\mathrm{IR} + e$$

If the risk-free rate of interest is 2%, what is the expected return on this portfolio?

Solution. The expected risk premium for this portfolio is 9%, as specified directly in the first term of the model. This represents the return in excess of the risk free rate, so the total expected return is 11%.

Notice that factor models are typically defined so that the factors reflect *unexpected* realizations of risk sources, such that their expected values are zero. Another way to say this is that whatever contribution the expected GDP growth and interest rate changes are expected to make to the total return is already reflected in the $E(R_i) = 9\%$ term in the model. In addition, the error term, e is defined to have a mean of zero.

Now, once the actual realizations for the GDP and IR factors are known, the extent to which they were above or below their expected values will, if the factor model is correct, impact the return on the portfolio by the product of the factor betas and the factor "surprise".

Question 2. What are the main assumptions underlying Arbitrage Pricing Theory?

Solution. The main assumptions are:

- Assumption 1 — Returns are generated by a factor model:

  $$R_i = E(R_i) + \beta_1 F_1 + \beta_2 F_2 + \cdots + \beta_N F_N + e_i$$

  This means that actual returns for any portfolio consist of its expected return, plus some return due to sensitivity to unexpected shocks that are common across all portfolios, plus an idiosyncratic risk term that is uncorrelated with the shocks.

- Assumption 2 — All assets can be bought or sold in any quantities at the market price, which leads to the key assumption that in equilibrium, arbitrage profits cannot be earned.

Question 3. Assume that a two factor APT model applies, that the expected return for a portfolio with sensitivity of 1.0 to factor 1 and 0.0 to factor 2 is 12% and the expected return for a portfolio with sensitivity of 0.0 to factor 1 and 1.0 to factor 2 is 6%. If the risk free rate is 5%, what is the expected return for a portfolio with factor 1 and factor 2 sensitivities of 0.5 and 1.5, respectively?

Solution. The basic form of a two factor APT model is easy to write when you are given the returns on the *factor portfolios* that have sensitivity of 1.0 to only one factor and sensitivity of 0.0 to all other factors, since these returns on the factor portfolios give you the risk premiums for the factors:

$$\begin{aligned} E(r_i) &= r_f + \beta_{i1}[E(r_1) - r_f] + \beta_{i2}[E(r_2) - r_f] \\ &= 5\% + \beta_{i1}(12\% - 5\%) + \beta_{i2}(6\% - 5\%) \end{aligned}$$

Then just plug in the betas for the portfolio in question:

$$\begin{aligned} E(r_i) &= 5\% + \beta_{i1}(12\% - 5\%) + \beta_{i2}(6\% - 5\%) \\ &= 5\% + .5(7\%) + 1.5(1\%) \\ &= 10\% \end{aligned}$$

Question 4. Suppose the market can be described by the following three source of systematic risk with associated risk premiums:

Factor	Risk Premium
Industrial Production (I)	6%
Interest Rates (R)	2%
Consumer Confidence (C)	4%

The return on a particular stock is generated according to the following equation:

$$r = 15\% + 1.0I + .5R + .75C + e$$

Find the equilibrium rate of return on this stock using the APT. The T-bill rate is 6%. Is the stock underpriced or overpriced?

Note that this is Question 11 from BKM Chapter 10 (Tenth Edition).

Solution. The formula given is a 3-factor model, which suggests that the expected return is 15% and that the actual return will also include the impact of shocks to I, R and C, as well as a firm-specific error term, e.

However, the question asked you to find the equilibrium return according to the APT, which states that the expected excess return should be given by the factor loadings and the factor risk premiums:

$$\begin{aligned} E(R) &= 1.0\mathrm{RP}_I + .5\mathrm{RP}_R + .75\mathrm{RP}_C \\ &= 1.0(6\%) + .5(2\%) + .75(4\%) \\ &= 10\% \end{aligned}$$

Given the risk-free rate of 6% and this 10% expected risk premium, the expected equilibrium return according to the APT is 16%. This is higher than the 15% based on the factor model, so we would say that the expected return is too low (relative to the returns on other assets with the same factor loadings) and therefore its price is too high.

Question 5. Suppose returns are generated by a one-factor model and that the risk free rate is 8%. Given the following betas and expected returns for two well-diversified portfolios, determine whether the market is in equilibrium or not and identify the arbitrage transaction to exploit any misplacing if it exists. Assume all of the assumptions of the APT model hold.

Portfolio	Expected Return	Beta
X	16%	1.00
Y	12%	0.25

Note that this is CFA Problem 2 from BKM Chapter 10 (Tenth Edition).

Solution. If returns are generated by a one-factor model, then according to the APT in equilibrium the returns on any well-diversified portfolio will be:

$$E(r_i) = r_f + \beta_1 \mathrm{RP}_1$$

Since we have been given the expected return for Portfolio X and we know the risk-free rate, we can use that to solve for the factor risk premium:

$$16\% = 8\% + 1.0\mathrm{RP}_1 \Rightarrow \mathrm{RP}_1 = 8\%$$

Using this, we can then calculate the equilibrium expected return for Portfolio Y:

$$E(r_Y) = 8\% + .25(8\%) = 10\%$$

Since the expected return is given as 12% and the equilibrium return is 10%, we would say that the market is not in equilibrium and that an arbitrage opportunity exists.

We could have reached this conclusion another way. According to the single factor APT, in equilibrium all well diversified portfolios have to have the same ratio of excess return to beta. In this case, that ratio for Portfolio X is 8% and for Portfolio Y it is 16%. From that alone, we know that we are not in equilibrium.

Notice that we can't actually say whether Portfolio X is mispriced or Portfolio Y is mispriced, only that they are mispriced relative to each other. Either way, we can exploit this arbitrage opportunity by buying a zero-investment portfolio such that the overall beta is zero.

Portfolio Y has a better expected return than the model would predict if Portfolio X were indeed priced correctly (it has a better return per unit of risk), so Portfolio Y is "cheap". To exploit an arbitrage, we buy what is cheap and sell what is expensive, so we would want to buy Portfolio Y and sell Portfolio X. To get a beta of zero, we need to buy four times as much of Portfolio Y as we sell of Portfolio X. To make it a zero investment portfolio, we would borrow the money we need for this.

For example, we could buy \$400 of Y, sell \$100 of X and borrow \$300 at the risk free rate. The expected return, in dollars, from this will be $400(12\%) - 100(16\%) - 300(8\%) = 8$ and it costs nothing. In addition, it is risk free because the factor sensitivities offset so that if the factor is 1% higher than expected, we will have $\$400(.25)(1\%) = \$1$ higher actual return from Portfolio Y and $-\$100(1)(1\%) = -\$1$ lower actual return from Portfolio X, which exactly offset each other.

Question 6. Returns are driven by a two-factor model and you are told that three well-diversified portfolios have expected returns and factor sensitivities as shown below:

Portfolio	E(r)	β_1	β_2
A	15%	0.80	1.20
B	14%	0.40	2.00
C	10%	0.24	0.40

Based on these portfolios you were able to estimate the following APT model:

$$E(r_i) = 7.75\% + 6.25\%\beta_{i1} + 1.875\%\beta_{i2}$$

Assume that a Portfolio E exists with factor sensitivities $\beta_1 = .46$ and $\beta_2 = 1.40$ and an expected return of 11%. Is there an arbitrage opportunity? If so, describe how you would exploit it.

Hint: Check to see what the factor sensitivities would be for a portfolio consisting of 25% in A, 50% in B and 25% in C.

Solution. First, plug the β_i's into the given APT model to find $E(r_E) = 13.25\%$. Since the expected return specified is only 11%, we know there is an arbitrage opportunity which involves finding another portfolio with the same sensitivities as this one and then buying the portfolio with the higher return and selling the portfolio with the lower return.

In this case, I hinted about a particular portfolio combining A, B and C. The factor sensitivities of this portfolio, call it D, are just the weighted averages of the respective factor sensitivities.

Therefore:

$$\begin{aligned}\beta_1 &= 25\%(.8) + 50\%(.4) + 25\%(.24)\\ &= .46\\ \beta_2 &= 25\%(1.2) + 50\%(2.0) + 25\%(.4)\\ &= 1.40\end{aligned}$$

Notice that this portfolio has the same sensitivities as Portfolio E, but it has an expected return equal to the weighted average expected return of its components, so:

$$E(r_D) = 25\%(15\%) + 50\%(14\%) + 25\%(10\%) = 13.25\%$$

So what would happen if we now bought \$100 if this Portfolio D and sold short \$100 of Portfolio E?

The first thing to notice is that it would not cost us anything now. So if it produced a positive expected profit with no sensitivity to any of the factors it would represent an arbitrage opportunity. What would be the sensitivities to the factors?

Here, $w_D = 100\%$ and $w_E = -100\%$, so we have the following weighted average betas:

$$\begin{aligned}\beta_1 &= 100\%(.46) + (-100\%)(.46) = 0\\ \beta_2 &= 100\%(1.4) + (-100\%)(1.4) = 0\end{aligned}$$

Therefore, this combination has no risk. What would we expect to earn from this? We'd get 13.25% return on our \$100 in D and we would pay 11% on our -\$100 in E, for a net return in dollars of \$13.25 − \$11 = \$2.25. This costs nothing up front, has no sensitivity to any of the factors and gives us \$2.25 next period. That is an arbitrage opportunity and we would exploit it as many times as we could.

Since that kind of opportunity cannot last for long, we know that when things settle down the two portfolios will have the same expected return.

Question 7. The following two portfolios have expected returns and factor sensitivities as shown below:

Portfolio	E(r)	β_1	β_2
A	10%	0.80	1.20
B	9%	0.40	2.00

Determine the parameters of an APT model of the expected return on a well-diversified portfolio if the risk free rate is 2.75%. Hint: This problem only requires basic algebra to solve.

Solution. We are looking for a model of the form:

$$E(R_i) = \beta_{i1}[\mathrm{RP}_1] + \beta_{i2}[\mathrm{RP}_2]$$

We are given two portfolios' expected returns and betas, so we can just plug in what we know and solve two equations for the two unknowns, RP_1 and RP_2. To simplify the notation, I will drop the '%' signs in the equations that follow and also convert the returns that were given into excess returns by subtracting the risk free rate.

$$7.25 = .8[\mathrm{RP}_1] + 1.2[\mathrm{RP}_2]$$

$$6.25 = .4[\mathrm{RP}_1] + 2.0[\mathrm{RP}_2]$$

Now, just write RP_2 as a function of RP_1 from the first equation:

$$\mathrm{RP}_2 = \frac{7.25 - .8\mathrm{RP}_1}{1.2} = 6.04 - .67\mathrm{RP}_1$$

Then plug that into the second equation and solve for RP_2:

$$6.25 = .4[\mathrm{RP}_1] + 2.0[6.04 - .67\mathrm{RP}_1] \Rightarrow \mathrm{RP}_1 = 6.25$$

And plus that in to find RP_2:

$$\mathrm{RP}_2 = 6.04 - .67\mathrm{RP}_1 = 1.875$$

Plugging these back into the original equation:

$$E(R_i) = \beta_{i1}[6.25] + \beta_{i2}[1.875]$$

And if we wanted to write this in terms of returns and not excess returns:

$$E(r_i) = 2.75\% + \beta_{i1}[6.25] + \beta_{i2}[1.875]$$

Question 8. In the BKM readings we used dominance arguments to show that certain assets with more attractive risk-return relationships might be preferred by some investors, depending on their risk preferences, and that investors would tilt their portfolios of risky assets towards holding more of what they perceived as being the underpriced (relative to their risk) assets. CAPM was based on this kind of argument.

We also used arbitrage arguments to show that when investors can identify investments with no net up-front investment but with risk-free profits, they will want to take large positions in those assets regardless of their risk preferences. APT was based on this kind of argument.

Describe why these two arguments result in different implications regarding the number of investors who are required to behave according to our model's assumptions in order to ensure that equilibrium prices achieve those in our models.

Solution. Dominance arguments result in each investor making small changes to their holdings, depending on their degree of risk aversion. To influence the prices of mispriced assets, we would need a large number of investors to be trying to optimize the mean and variance of their portfolios so that the aggregate pressure on prices was sufficient to bring about equilibrium.

In contrast, arbitrage arguments require only *one* investor willing and able to identify the arbitrage opportunity and then mobilize large dollar amounts (recall that in APT we assumed unlimited ability to buy or sell any asset) in order to restore equilibrium.

Question 9. You have estimated the factor risk premiums in an APT model and have used unexpected inflation as one of the factors in the model. Your analysis shows that the factor risk premium on inflation is negative. Does this mean that your model assumes that portfolios with inflation risk have lower expected returns than stocks that do not have inflation risk?

Solution. No, it doesn't. It says that portfolios with positive betas (positive factor loadings) on the inflation risk factor will have lower expected returns.

But what kinds of portfolios will have positive factor loadings on a factor measured as the amount of unexpected inflation each period? They will be portfolios whose returns are high when unexpected inflation is high. Those are *desirable* portfolios because they have higher payoffs at bad times. Investors would be willing to pay more for these portfolios because they hedge inflation risk. Paying more means that they will have lower expected returns.

As this example shows, the sign of the factor risk premiums is a function of whether exposure to the factor is desirable or not. We should expect the risk premiums to be positive if it is a risk factor you do not want exposure and negative if it is a risk factor you do want exposure to.

For most portfolios, the factor loadings on this inflation factor are likely to be negative. That is, most portfolios will have negative returns at times when unexpected inflation is high. This will then give the reasonable sign for the inflation term (factor loading times factor risk premium), which is that for most portfolios the existence of inflation risk will mean a positive expected risk premium.

BKM Chapter 11: Market Efficiency

This chapter is not very technical but it is loaded with important information. The textbook should be read carefully and you can refer to the notes below for the highlights.

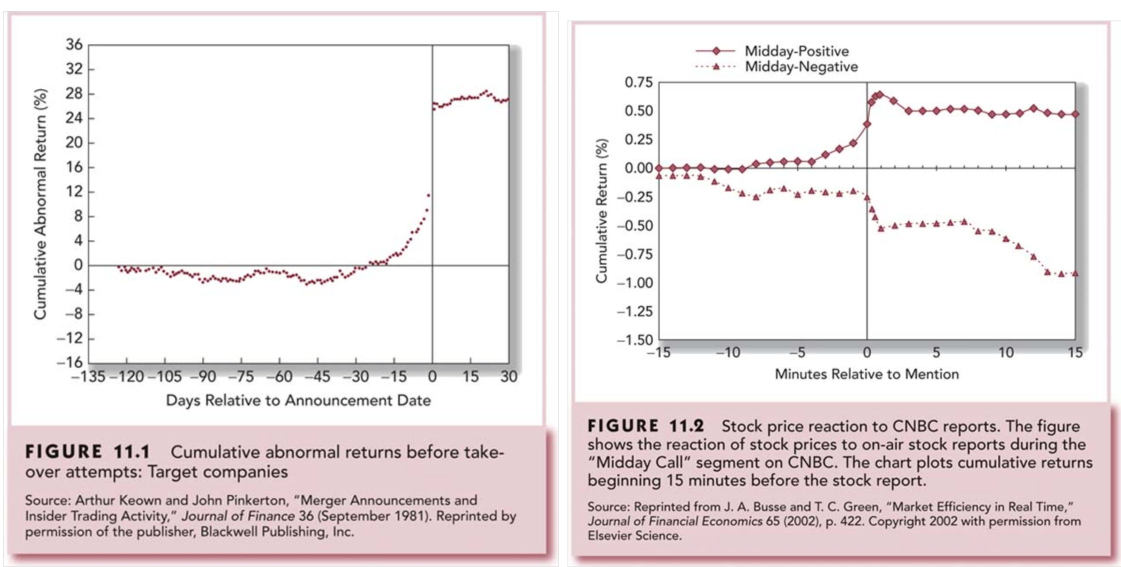
Random Walk

Studies of past stock prices tend to support the notion that stock price changes are random and unpredictable. When stock prices behave in this fashion, we refer to this as a random walk. Note that we are not saying that prices themselves are random, only that the price changes are random — for instance because the changes reflect the recognition of new information which arrives in an unpredictable way.

Efficient Market Hypothesis (EMH)

Underlying most of modern finance is the notion that if a market is efficient then prices will reflect all currently available information. In such a market, prices will respond quickly (though not necessarily correctly!) to new information. Ample evidence of this exists, including news of takeovers or analysis of firms featured in CNBC reports. These two examples are shown in BKM Figures 11.1 and 11.2, respectively, which are reproduced as Figure 1.

FIGURE 1. Examples of Market Response to Information



Of course, there's no reason to think that all markets are necessarily efficient — or even that any of them are. But competition to discover what other investors have overlooked, and the ample rewards that could be obtained, suggest that a lot of people will indeed be trying to gain an edge that would result in information being quickly incorporated.

There are three versions of the efficient market hypothesis:

- **Weak Form** — Current prices reflect all information that can be obtained from examining past price history, trading volume, etc. All of this information is easily available and virtually costless to obtain, so if this provided any insight into future prices it would be reflected in the current prices.

- **Semi-Strong Form** — All publicly available information regarding the prospects for the firm is reflected in the stock price.

- **Strong Form** — All information, including information known only by company insiders, is reflected in the current prices.

Technical Analysis

This type of analysis involves extensive examination of past stock prices in the search for recurring and predictable patterns. Since this information is so readily available, we should expect intense competition to uncover patterns to eliminate trends the moment they become evident. So even if technical analysis were capable of discovering trends, it ought to quickly become self destructing as people attempt to exploit the information.

People who use these methods are often referred to as chartists because of their focus on stock price charts.

We will examine some of these methods in more detail in the next chapter.

Fundamental Analysis

Rather than studying past prices, fundamental analysts use valuation methods such as discounted cash flow to constantly value stocks on a ground-up basis. By carefully forecasting the earnings and dividend prospects for the firm and taking full consideration of all available information, they hope to uncover instances where stock prices deviate from this fundamental value and to exploit those differences.

Because so many other investors are similarly trying to determine the fair value for the stocks and exploiting any differences, successful fundamental analysts must be more insightful than other investors.

Different Points of View Regarding Value

Notice that technical analysts and fundamental analysts essentially adopt different views regarding what *value* represents. Technical analysts believe that an asset is worth whatever someone will pay for it and so they focus on trying to predict what others will pay for assets in the future. They will buy assets that they believe others will value more highly in the future, regardless of the reason.

In contrast, fundamental analysts believe that assets have intrinsic value and merely attempt to discover what this value is. They will buy assets when they can pay less than this intrinsic value for them.

Active vs. Passive Strategies

Neither technical analysis nor fundamental analysis is likely to be of much value to most investors since so many others are using similar methods to also search for and exploit mis-priced securities. Only those investors who possess unique insights or information should be able to exploit mispriced securities in an efficient market, suggesting that for most investors the effort and cost involved will probably overwhelm the payoff from doing the analysis in the first place. For many investors, a passive strategy such as holding index funds is probably the best strategy.

Why Bother at All?

If stock markets are efficient and passive strategies are preferred, why bother with any portfolio management? The EMH simply suggests that you are unlikely to "beat the market"; you still have to be concerned with selecting a diversified portfolio that reflects your specific degree of desired risk. In addition, factors such as your age, tax bracket and employment must also be reflected.

Event Studies

One consequence of the EMH, if it holds, is that market prices can be used to study the impact of new information, such as the market's assessment of the value of an acquisition after it is announced. This process is known as an event study.

Recall that if we assume a single index model, then the return for a stock can be written as the sum of a fixed component, a market component and a firm specific component:

$$r_t = a + br_M + e_t$$

The firm specific component represents the return beyond what would have been expected given the market movement alone, or the abnormal return.

Then an event study simply examines the returns on the day of a given event, determines the abnormal return and then attributes the abnormal return to the specified event. Usually, several days prior to and after the event date are studied and the cumulative abnormal returns are measured to account for the potential leakage of information to the market prior to the official event date (e.g. rumors of a takeover before it is announced) and to also reflect the potential for the market to be slow to fully reflect the impact of the event.

Cumulative Abnormal Returns Surrounding Takeover Announcements

The text gives some event study examples. One looks at the cumulative abnormal returns around the takeover announcement dates and shows that abnormal returns begin to occur about 15 days prior to the announcement but that on the date of the announcement there tends to be a large abnormal return, followed by no abnormal returns. This shows that the market immediately reflects the value of the information when it is announced.

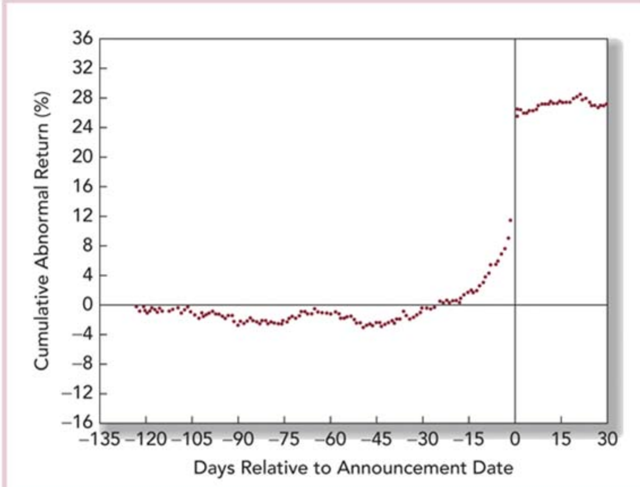


FIGURE 11.1 Cumulative abnormal returns before takeover attempts: Target companies

Source: Arthur Keown and John Pinkerton, "Merger Announcements and Insider Trading Activity," *Journal of Finance* 36 (September 1981). Reprinted by permission of the publisher, Blackwell Publishing, Inc.

Are Markets Efficient?

Any market could exhibit different degrees of efficiency. By definition, an efficient market is one in which intense competition for an informational advantage causes all publicly available information to be immediately reflected in prices. So the real point is that you ought to think of markets in terms of their likelihood to exhibit these qualities. If they do, then prices likely reflect all of this information.

But there is an endless debate over the issue of whether the U.S. stock market is efficient. If it is, then it will not be possible to earn excess risk adjusted returns consistently. Note the subtle but important point that the EMH does not say that people cannot earn profits from investing, nor does it say that they cannot earn more than the market return — it only says that they cannot earn returns that exceed those that should be expected given the amount of risk they take.

Unfortunately, there are three key issues that prevent us from ever settling this definitively:

- *Magnitude Issue* — Even if the markets are very nearly efficient, a small degree of inefficiency could still lead to substantial profits to those who are able to exploit it. But given the statistical noise, measuring this tiny degree of inefficiency is likely impossible.

- *Selection Bias* — It is impossible to know with certainty whether there exist strategies to beat the market because those who possess such abilities are unlikely to want to publicize this. Since we will only hear about the strategies that do not work, we can never know for certain if other strategies do exist.

- *Lucky Event Issue* — Even if we did identify people who could beat the market on a risk adjusted basis, there's no way of knowing whether this is just luck or some sort of repeatable outcome.

Tests of Weak Form Market Efficiency

Early tests of the EMH were tests of whether past stock prices could be used to predict future prices. That is to say, are there patterns in stock prices (such as those that technical analysts seek to identify) that can be exploited to earn abnormal returns?

Returns Over Short Horizons

Tests of short-term serial correlation in broad market indices suggest that there is some positive correlation of returns from week to week (or even month to month). In addition, at the individual stock level there seems to be negative serial correlation. However, the effects in both cases are too small to suggest any ability to profit from such correlations after taking into account transaction costs.

Although individual stocks do not exhibit substantial serial correlation, over the intermediate term portfolios tend to exhibit positive serial correlation. There's a *momentum effect* that tends to cause portfolios of the best and worst performers (over the past 3–12 months) to continue their performance (over the subsequent 3–12 months, but with some reversal in the 2 years following that).

Returns Over Long Horizons

There is some evidence that in the long run stock returns (for the overall market) are negatively correlated, exhibiting *mean reversion*, perhaps caused by overreaction to news events and then subsequent correction. However, this fad explanation need not be what is going on—the negative serial correlation could also be caused by changes in the risk premium or overall discount rates.

The same overreaction and reversal pattern found for the overall market seems to exist for individual stocks too. The best performers in the past 3–5 years tend to underperform in the future (over the subsequent 3–5 years) and the worst performers tend to outperform in the future. This evidence suggests that a contrarian strategy of buying losers and selling winners should be profitable.

Broad Market Predictors

Several studies have found that variables such as dividend yields, earnings yields, bond spreads, can be used to predict broad market returns. However, this is more likely simply predicting changes in risk premiums rather than providing evidence of stock market predictability. For example the yield spread on bonds is a better predictor of stock returns than bond returns; the dividend yield is also a good predictor of bond returns. Both of these suggest that these variables are related to risk premiums that affect both markets.

Semi-strong Tests and Market Anomalies

If markets are efficient, then we should not be able to find simple trading rules based on publicly available information that lead to excess risk adjusted returns. However, researchers have found a number of examples of so-called *anomalies* that seem to defy the EMH.

Before summarizing the main points, note though that in these examples it is necessary to *risk-adjust* the returns and therefore we have to use a model such as CAPM or APT to determine what the risk adjusted returns should have been. But if our risk adjustment is incorrect, it could mistakenly suggest that an excess return exists when in fact it is just a failure of our risk adjustment.

P/E Effect

Evidence suggests that low P/E stocks outperform high P/E stocks, even after adjusting for differences in betas between the various groups. This would be contrary to the EMH because ranking stocks based on their P/E ratio is so easy that we would expect people to do this without much effort and then force up the price of low P/E stocks and force down the price of high P/E stocks. But in reality, perhaps beta is simply not fully reflecting the risk differences in the low P/E and high P/E stocks.

Small Firm in January Effect

Banz showed that ranking firms by their total market capitalization we can see that small firms tend to significantly outperform large firms. Again, even after adjusting for differences in betas, this small firm effect remains and is rather significant. This too would contradict the EMH since such a simple rule should not be able to result in such excess profits.

It turns out that virtually all of the small firm effect occurs in January, and usually in the first few days. One explanation of this is the impact of tax loss selling which tends to cause mostly small stocks to be sold in December to lock in tax write-offs and then are repurchased in early January.

Neglected Firm Effect

Researchers found that firms that are followed less by analysts (e.g. smaller firms) tend to have less readily available information and tend to earn excess risk adjusted returns. This is likely because the lack of readily available information makes these firms riskier in ways that the CAPM beta itself does not capture.

Recall from Chapter 9 that illiquidity effects can cause the prices of smaller, neglected firms to be significantly reduced, resulting in a rate of return premium. This could partially explain both the small firm in January effect and the neglected firm effect. However, it wouldn't explain why the small firm returns should be high only in January. Nonetheless, the high trading costs of illiquid small and neglected firms could wipe out any apparent opportunities to earn abnormal returns.

Book to Market Ratios

As discussed in the notes for Chapter 10, Fama and French have found that firms with the highest ratios of book value to total market capitalization tend to outperform firms with the lowest ratios. More importantly, they found that once both the size and the book to market effects are taken into account, the standard CAPM beta seems to have no power to explain average returns.

Post Earnings Announcement Price Drift

Often there is a sluggish response to earnings announcements. In one study, analyst estimates of earnings were compared to actual earnings to estimate the earnings surprise and then 10 portfolios were formed with the smallest to largest surprises. The abnormal returns for each portfolio were then compared for the 60 days before and after the earnings announcement.

For firms with positive surprises, the cumulative abnormal returns are positive on the announcement day. For firms with negative surprises, the CAR is negative. However, the CARs tend to drift further upwards or further downwards after the announcement date. This suggests that people don't fully reflect the surprises fast enough and that a simple strategy of buying the positive surprise firms and selling the negative surprise firms would have been profitable.

Strong Form Tests: Inside Information

Several studies have documented that insiders do in fact earn abnormal returns, which is really no surprise. However, once the insider trades become public, it is not possible to earn abnormal returns from this information.

Risk Premiums or Inefficiencies?

The anomalies mentioned above are quite puzzling. Of course, they are all somewhat related because stocks suffering a large drop in price would show up in many of these categories (small firm, high book to market ratio, relative loser). Second, these variables may simply be proxies for some factor in an APT-type model, suggesting the excess returns are simply risk premiums that are entirely consistent with an efficient market. This is the argument that Fama and French make with regard to the results of their 3-Factor model.

On the other hand, others have argued that these anomalies present evidence of systematic mispricing of stocks, such as what would result if analysts made systematic errors in their forecasts. For instance, if analysts mistakenly extrapolate recent performance too far into the future then prices will overshoot and exhibit reversal patterns. But firms with sharp drops will also wind up being smaller firms or firms with high book to market ratios. As a result, the Fama-French results could indeed reflect market inefficiencies rather than true risk premiums.

Anomalies or Data Mining?

One easy criticism of all of these findings of anomalies is that it is too easy to examine data until anomalies are found but an altogether different exercise to predict anomalies before they occur. Many of the anomalies discussed tended to disappear shortly after being discovered (e.g. small firm effect in the 1980's, book-to-market effect in the 1990's). Nonetheless, some of these anomalies represent real puzzles that are yet to be explained, such as why value stocks have tended to provide higher returns than growth or glamour stocks. Because similar size, momentum and book-to-market effects seem to appear in several security markets around the world, they most likely represent unexplained risk premiums (my opinion), but this is far from a settled matter.

Bubbles and Market Efficiency

If markets are broadly "efficient", why do prices sometimes seem (in retrospect at least) to lose their grounding in reality—such as during the dot-com bubble or in the housing price bubble that led to the 2008 financial crisis?

One argument is that periods of stability cause investors to become more willing to assume risk. Declining risk premiums result in rising prices, which leads to more optimism and even lower required risk premiums. The challenge though is to spot bubbles in advance, a feat made more difficult by the fact that often "bubble" have some basis in reality and only become obvious signs of mispricing after the fact.

A bigger issue is the difficulty and risks associated with trying to take advantage of instances of mispricing—short selling could be costly, locating securities to sell short could be difficult and mis-timing the market's price corrections could lead to substantial losses in the short-term.

Mutual Fund and Analyst Performance

If markets are efficient, then consistent abnormal risk adjusted trading profits should not be achievable, even for professional investors. The text examines the performance of stock analysts who make investment recommendations and the performance of mutual fund managers who actually manage portfolios.

Stock Market Analysts

Interpreting analyst recommendations is particularly challenging because most of them worked for brokerage firms and they tended to be overwhelmingly positive. However, if we study changes in recommendations, we see that positive changes are associated with 5% stock price increases and negative changes with 11% decreases. Noting that this could be due to the impact the recommendations themselves have on prices, researchers believe that this reflects analyst skill (at least in part). Other researchers have looked at consensus recommendations and found some abnormal returns, but note that transaction costs would wipe out potential profits.

Mutual Fund Managers

The raw evidence shows that for the most part mutual fund managers do not beat a passive index in most years. However, some managers do beat the market, but not on a risk adjusted basis. Malkiel looked at returns for mutual funds and found that the alphas were essentially normally distributed with a mean of approximately zero.

Malkiel's study used the S&P 500 as the market portfolio. This is somewhat inappropriate because a lot of mutual funds hold small stocks and other non-S&P 500 assets and deviations from market returns could be due more to this style choice rather than to any stock picking ability. Instead, it may be better to use a multi-factor model with the excess returns on the S&P 500, a small stock index and a bond index as the three factors. When this is done, mutual funds seem unable to beat a passive strategy with comparable risk characteristics. Even more interesting, the results are worse for funds with higher expense ratios—it seems the extra money spent searching for superior returns does not lead to high enough returns to justify the expense.

A similar analysis by Carhart that uses the Fama-French 3-Factor model with a fourth factor representing momentum allows for a wide range of mutual fund styles to be reflected and finds that alphas are generally negative. With regard to persistency, he finds only minor persistence in relative performance, most of which appears to be due to expenses and transaction costs rather than gross investment returns.

Others find persistence over very short horizons, but find that the top performers' outperformance one quarter drops substantially in the next quarter and that the persistence is too small to justify performance chasing. This supports the idea that the best performers will attract more investment funds, driving the alpha towards zero.

Similar results were obtained from examinations of bond fund managers.

Practice Questions

Question 1. Describe the concept of market efficiency, including the three major forms.

Solution. If a market is efficient then prices will reflect all currently available information. There are three versions of this:

i. Weak Form — Current prices reflect all information that can be obtained from examining past price history, trading volume, etc. All of this information is easily available and virtually costless to obtain, so if this provided any insight into future prices it would be reflected in the current prices.

ii. Semi-Strong Form — All publicly available information regarding the prospects for the firm is reflected in the stock price.

iii. Strong Form — All information, including information known only by company insiders, is reflected in the current prices.

Question 2. Describe how both technical analysis and fundamental analysis can affect market efficiency.

Solution. Technical Analysis involves the study of past stock prices in the search for recurring and predictable patterns. Since past price information is readily available, intense competition to uncover patterns will likely eliminate the trends the moment they become evident. In other words, if technical analysis were capable of discovering trends, people would quickly attempt to exploit the trends by either buying before the price rises or selling before the price falls. This will eliminate the lag from the start of the trend to the final resolution and effectively eliminate the pattern that was being detected. In the end, patterns will not be readily visible and the market will appear to be efficient in the sense that prices will reflect the information content (whatever it may be) in past prices.

Fundamental Analysis involves methods to value stocks on a ground-up basis. By carefully forecasting the earnings, cash flow and dividend prospects for the firm and taking full consideration of all available information, fundamental analysts hope to uncover instances where stock prices deviate from this fundamental value and to exploit those differences. Because investors using this approach are always trying to identify underpriced or overpriced stocks, buying the underpriced ones and selling the overpriced ones, at any point in time it will be quite difficult to find obvious examples of underpriced or overpriced securities. This will make the market appear efficient in the sense that prices will reflect relevant information available.

Question 3. List and briefly describe five market anomalies.

Solution. Five that are discussed in the text include the following:

- Momentum — Tests of weak form market efficiency tended to focus on the degree of serial correlation in stock prices. Small week to week or month to month correlations were suggested for individual stocks, but these were too small to exploit profitably. However, portfolios were found to have larger intermediate term serial correlation. This suggests a momentum effect that tends to cause portfolios of the best and worst performers to continue their performance from the recent past (best and worst performers over the past 3–12 months tend to also overperform and underperform in the subsequent 3–12 months).

- P/E Effect — Evidence suggests that low P/E stocks outperform high P/E stocks, even after adjusting for differences in betas between the various groups. This would be contrary to the EMH because ranking stocks based on their P/E ratio is so easy that we would expect people to do this without much effort and then force up the price of low P/E stocks and force down the price of high P/E stocks. But in reality, perhaps beta doesn't fully reflect the risk differences in the low P/E and high P/E stocks.

- Small Firm Effect (in January) — Banz showed that by ranking firms by their total market capitalization we can see that small firms tend to significantly outperform large firms. Even after adjusting for differences in betas, this small firm effect remains and is rather significant. This too would contradict the EMH since such a simple rule should not be able to result in such excess profits. But it turns out that virtually all of the small firm effect occurs in January, and usually in the first few days. One explanation of this is the impact of tax loss selling which tends to cause mostly small stocks to be sold in December to lock in tax write-offs and then are repurchased in early January.

- Neglected Firm Effect — Researchers found that firms that are followed less by analysts (e.g. smaller firms) tend to have less readily available information and tend to earn excess risk adjusted returns.

- Book to Market Ratios — Researchers have found that firms with the highest ratios of book value to total market capitalization tend to outperform firms with the lowest ratios.

Question 4. Describe what is meant by Post Earnings Announcement Price Drift.

Solution. In one study, analyst estimates of earnings were compared to actual earnings to estimate the earnings surprise. Ten portfolios were formed with groups of stocks with the smallest to largest surprises. The abnormal returns for each portfolio were then compared for the 60 days before and after the earnings announcement. For firms with positive surprises, the cumulative abnormal returns (CAR) are positive on the announcement day. For firms with negative surprises, the CAR is negative.

FIGURE 2. BKM Figure 11.5

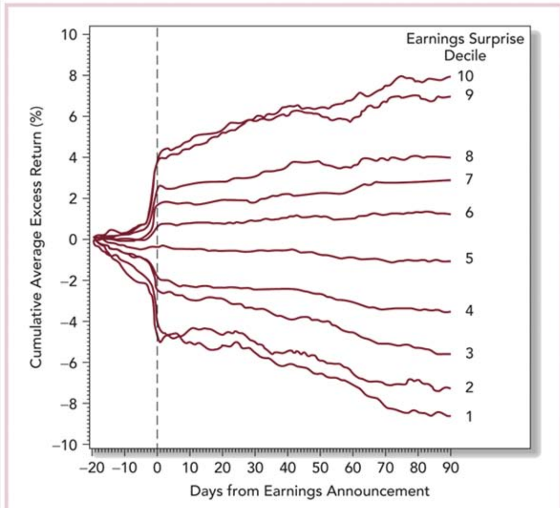


FIGURE 11.5 Cumulative abnormal returns in response to earnings announcements

Source: Reprinted from R.J. Rendleman Jr., C. P. Jones, and H. A. Latané, "Empirical Anomalies Based on Unexpected Earnings and the Importance of Risk Adjustments," *Journal of Financial Economics* 10 (1982), pp. 269–287. Copyright 1982 with permission from Elsevier Science.

This alone suggests market efficiency, since the prices seemed to reflect expected results and reacted accordingly when those expectations proved to be too high or too low.

However, the CARs tend to drift further upwards or further downwards after the announcement date. This suggests that people don't fully reflect the surprises quickly enough in the price. This phenomenon is referred to as Post Earnings Announcement Drift. A simple strategy of buying the positive surprise firms and selling the negative surprise firms would have been profitable.

Question 5. What is an event study?

Solution. One consequence of the EMH, if it holds, is that market prices can be used to study the impact of new information, such as the market's assessment of the value of an acquisition after it is announced. This process is known as an event study.

An event study simply examines the returns on the day of a given event, determines the *abnormal return* relative to a model such as CAPM or a single-index model and then attributes the abnormal return to the specified event.

Usually, several days prior to and after the event date are studied and the cumulative abnormal returns are measured to account for the potential leakage of information to the market prior to the official event date (e.g. rumors of a takeover before it is announced) and to also reflect the potential for the market to be slow to fully reflect the impact of the event.

Question 6. You have collected return data for a 40 year period for portfolios constructed using the top one-third (biggest) and the bottom one-third (smallest) stocks in the U.S. stock market. The table below shows the average annual excess returns (over and above the risk-free rate) for these portfolios, along with the historical average excess return for the entire market for the same period. Also shown in the table is the average market beta for the two portfolios.

	Avg Excess Return	Avg Market Beta
Small Capitalization Stocks	8.7%	1.02
Big Capitalization Stocks	6.3%	1.01
Market Risk Premium	6.2%	

Based on the data above, determine whether this data suggests an anomaly with respect to the semi-strong form of the efficient market hypothesis and the Capital Asset Pricing Model.

Solution. The two portfolios suggest that smaller firms (as measured by their market capitalizations) have higher average excess returns than bigger firms. If their market betas were different from each other, this would not necessarily be an indication of an anomaly since the return differential could just be compensation for market risk.

In this case, the betas are quite close (1.02 vs. 1.01) and this small difference, when combined with the market risk premium shown of 6.2%, cannot account for the large difference in returns. Using the CAPM as the reference point, we would therefore conclude that excess returns unrelated to market risk could be achieved through a simple sorting algorithm – sort companies by size and buy the small firms/sell the large firms – that would capture this premium without any market risk.

Of course, this perceived source of excess return could simply reflect compensation for some other type of risk not captured in the CAPM. This is, in fact, one of the additional risk factors used in the Fama-French model. But if the CAPM were the reference point, these results would appear anomalous with the semi-strong form of the efficient market hypothesis.

Note that the 2014 exam contained a question (6a) that was trying to get at this same point. However, they did not give you the market betas for the portfolios and, in the absence of those betas, it is impossible to say whether the returns are consistent or not with the CAPM. The question also did not mention what asset pricing model to use as a reference point, so broad statements about market inefficiency would also be suspect. All such statements must be in reference to, or relative to, a well-defined asset pricing model using risk-adjusted returns.

BKM Chapter 12: Behavioral Finance and Technical Analysis

The Behavioral Critique

The premise of behavioral finance is that conventional financial theory ignores how people really make decisions under uncertainty. First, there is experimental evidence that suggests that people commonly make errors inferring probability distributions. Second, even when they know the probability distributions, they often make systematically suboptimal decisions.

As we will see, these are interesting observations, but their impact on security markets is less clear. If these sorts of errors lead to mispriced securities, then we should expect arbitrageurs to take advantage of these errors. Their actions would eventually cause the errors to be corrected. So if these errors do impact market prices, there must be some other limits to the actions of the arbitrageurs.

Information Processing Errors

Errors in information processing can cause investors to misestimate probability distributions. The following are some examples of these types of errors:

- Forecasting Errors — People tend to give too much weight to recent evidence and tend to produce extreme forecasts. This could lead to excessive earnings forecasts and cause high P/E stocks to subsequently under-perform.

- Overconfidence — People tend to exhibit extreme overconfidence and overestimate their abilities.

- Conservatism — Investors are often slow to update their prior beliefs in the face of new information.

- Sample Size Neglect and Representativeness — People tend to infer patterns from limited information.

Behavioral Biases

Even with perfect information, investors often make irrational decisions that lead to suboptimal risk-return trade-offs. Here are several examples of the things people often do to cause this:

- Framing — Decisions tend to be influenced by the way they are framed. For instance, people tend to be risk averse when it comes to gains (would prefer a certain gain to a gamble) but are risk seeking when it comes to losses (would prefer a gamble to a certain loss).

- Mental Accounting — A specific type of framing is when people segregate certain decisions. For example, some gamblers may make different decisions when playing with the house's money versus when they have to dip into their own funds.

- Regret Avoidance — People tend to blame themselves less when their choices are more conventional, which causes people to shy away from out of favor stocks. This could explain why higher returns are required to entice investors to buy firms with high book to market ratios.

- Prospect Theory — Recall from Chapter 6 that standard financial theory relies on the assumption that risk averse investors evaluate risky propositions by attempting to maximize their utility. Further, utility functions commonly assume that utility is a function of wealth and that people become less risk averse as their wealth increases. In contrast, *prospect theory* assumes that people are loss averse and that they evaluate utility based on *changes* in wealth. Since they continually re-center what they perceive as gains and losses, they do not become less risk averse as their wealth rises. Further, their utility functions are convex with respect to losses and therefore they tend to be risk-seeking rather than risk averse when it comes to losses.

Limits to Arbitrage

Recall that the key argument to why we would expect many markets to be efficient, even if some people get it wrong, is that there are plenty of other people searching for information that will give them an advantage and they are always ready to exploit any advantage they find. But several factors limit the ability of others to profit from mispricings:

- Fundamental Risk — Suppose you are confident a stock is underpriced. The argument is that you should buy it and wait for the mispricing to go away. Your actions (and those of others who do the same thing) will drive the price up and eliminate the underpricing that you identified.

  But what if the underpricing actually gets worse before it gets better? Can you afford to wait forever? If you leveraged your bet with borrowed funds so that you could really exploit the mispricing, your debt could come due before the price rises (e.g. the case of Long Term Capital Management). Or if you invest on behalf of clients, they could fire you for the short term losses before the price rises. So even if you are right, the strategy is far from risk free for you.

- Implementation Costs — Often mispricings cannot be fully exploited due to transaction costs or other limitations.

- Model Risk — What if you are wrong? All opportunities to exploit identified mispricings will carry some risk and therefore may limit the willingness to try to exploit it fully.

Limits to Arbitrage and the Law of One Price

The text offers three examples where it seems as though the basic idea that two identical assets should have the same price (the Law of One Price) has been violated:

- Siamese Twin Companies — After Royal Dutch and Shell merged, the two separate stocks continued to trade but the firms agreed to a 60/40 split of all profits of the combined company for Royal Dutch and Shell, respectively. As a result, it would seem as though Royal Dutch's stock should always trade for 1.5 times that of Shell, otherwise there would be an arbitrage profit. However, over long periods of time this was actually not the case, with deviations ranging from more than -5% to 17%. Investors were not always willing to take advantage of this arbitrage opportunity due to the fundamental risk involved — over short periods of time the mispricing could get worse, posing short term mark to market losses with potentially serious consequences.

- Equity Carve Outs — An equity carve out occurs when a company sells an equity stake in a subsidiary or division. In some cases, only a small portion of the shares will be traded, with the rest of the shares given, at some later date, to the existing shareholders. During the technology bubble of the late 1990's, 3Com did this with their Palm subsidiary by selling 5% of Palm to new investors and arranging for 3Com's existing shareholders to receive the other 95% of Palm's shares (specifically, they were to receive 1.5 shares of Palm for each share of 3Com they owned).

  As a result, 3Com's price should have been at least 1.5 times the price of Palm prior to the distribution of the Palm shares to 3Com's shareholders, plus the value of all of 3Com's other businesses. However, for an extended period of time, 3Com's stock price was actually less than the price of Palm.

  Apparently investors were overvaluing Palm and/or undervaluing 3Com, making it possible for an arbitrage profit by purchasing 3Com and shorting Palm. However, it was actually quite difficult and expensive to short Palm at the time, due to the very small volume of Palm shares outstanding and the fact that they were already sold short.

- Closed-End Funds — With open-ended mutual funds, investors are always able to redeem their shares directly with the fund manager at whatever the market value of those shares are. In contrast, closed-end funds do not allow redemptions. Instead, investors seeking to sell their shares simply sell them to other investors, just as they would common stock. Interestingly though, the market value of the mutual fund shares often reflects a substantial discount from the net asset value (NAV), which suggests that if the fund were to simply sell all of its holdings, the investors would receive a windfall since the holdings would be sold for the NAV.

  This is somewhat of a puzzle, though to some extent deviations from NAV are to be expected since the fund does incur expenses (suggesting discounts to NAV might be reasonable) and the fund managers may indeed be able to earn a positive risk-adjusted return (suggesting a premium to NAV might be reasonable).

However, researchers have found that the discounts on different funds tend to move together and are correlated with small stock returns, suggesting that all might be impacted by variations in investor sentiment. In addition, while it may be theoretically plausible to buy the funds trading at a discount and sell (short) the funds trading at a premium, eventually earning an arbitrage profit when prices converge to NAV, this would carry significant fundamental risk since the premiums and discounts could widen.

It is also important to note that this anomaly of closed-end funds trading at a discount or premium might indeed be perfectly rational. It turns out that the premium (discount) is a simple function of the dividend yield, the expenses and the alpha the fund managers can earn. Even if the expenses and alpha are small, a modest alpha in excess of the expenses could result in a sizable premium; expenses in excess of the alpha could lead to a significant discount as well. When a fund's shares are first issues, it might be natural to assume that investors believe the fund can earn a positive alpha, hence the premium to NAV that is often found at issue. However, once the fund fails to deliver positive alpha, this could lead to a significant discount to NAV.

Bubbles and Behavioral Economics

In the late 1990's, prices for technology stocks in particular and the Nasdaq index more generally (which is heavily weighted towards technology stocks) rose significantly (6-fold in the 5 years beginning in 1995). Prices seemed to get way out of line with *fundamental value*, driven by what appeared to be irrational exuberance, and ultimately prices by October 2002 were one-fourth their previous high.

This episode caused many to question how efficient the market could be with the broad market exhibiting all of the behavioral patterns noted earlier — irrational investor sentiment driving prices, overconfidence of investors, representativeness bias as investors used short-term patterns to forecast the distant future. Indeed, one study found that firms enjoyed substantial increases in market value by simply adding ".com" to their corporate name, without any changes to their underlying business.

In fact, it is very hard to tell a bubble is occurring until after it has popped. Recall that during the technology bubble many commentators argued that we were in a new economy and the full potential for the internet and technology to transform the economy was not well understood. Indeed, rather modest errors in projecting future growth rates can cause substantial differences in estimates of the value for the overall market.

Behavioral Interpretations of the Fama-French Model Empirical Results

In Chapter 10 we discussed the Fama-French model attempted to explain the return differentials for small vs. large and high book-to-market vs. low book-to-market firms unrelated to differences in their betas.

Fama and French interpreted these results as suggesting that theirs SMB and HML factors proxy for priced risk factors. With that interpretation, CAPM as a model of returns is seriously questioned (which is why almost all research done now relies on the Fama-French model and not the Standard CAPM), but the ideas that markets are efficient and that only systematic risk factors are priced both valid.

But another way to interpret the evidence that the HML portfolios help to explain returns on stocks, thus resulting in a *value premium*, is that investors may irrationally prefer certain glamour firms that have had good recent performance, high prices and thus lower book-to-market ratios. This causes their returns to lag behind the value firms.

The behavioral rationale for this might be due to the representativeness errors discussed earlier. These errors cause investors to extrapolate recent performance too far into the future and to overreact to good news. The evidence that this is the case is that when firms are ranked by their past growth rates in income, their book-to-market ratios decline steadily. This suggests that investors extrapolate high growth too far into the future, overvalue the firm and cause the book-to-market ratios to fall. As a result, high book to market firms will tend to be those whose prices were not exaggerated and their returns will, on average, be higher.

Notice that if low book-to-market firms (the glamour firms) did indeed have higher growth, investors would be right to value them highly. However, when the low book-to-market firms are identified up-front, their future five year growth rates are not higher than other firms. That is, it does seem as though the book-to-market ratios reflect past rather than future growth.

Additional evidence comes from an analysis of earnings announcements for firms classified as either growth (low book-to-market ratio) or value stocks (high book-to-market ratio). On average, investors are relatively disappointed with the earnings announcements of growth stocks. It seems as though they optimistically bid up the price of these stocks (causing their book-to-market ratios to be low) based on expectations of good earnings and were generally disappointed when actual earnings were announced.

Evaluating the Behavioral Critique

Efficient market theory suggests that market prices are to a large extent reliable. The behavioral critique calls this into question, but it fails to offer much evidence that there are exploitable patterns. Yes, investors are far less rational than we often assume, but does this lead to behavior and to market prices that we can exploit to earn abnormal risk-adjusted returns?

The behavioral biases mentioned earlier offer a laundry list of interesting behaviors that can be used to explain just about any anomaly. This isn't all that useful on its own without a consistent or unified behavioral theory.

Technical Analysis

The concept of technical analysis, whereby the analyst attempts to exploit predictable patterns in stock prices, was mentioned briefly in the previous chapter. When evaluating all trading rules that result from searching for patterns, it is important to note that most patterns can only be detected after-the-fact. Indeed, often completely random price movements can lead to seemingly detectable patterns whose future path is completely unpredictable.

Nonetheless, some of the main tools of technical analysts are discussed in the textbook, in part because of the growing recognition that the behavioral finance issues discussed earlier could indeed lead to some of the observed patterns and efficient market anomalies.

Do Behavioral Biases Lead to Exploitable Patterns?

The disposition effect suggests that investors don't like to sell losing investments. This will cause the demand for stocks to depend on the price history of the shares (when each of the investors bought the stock) and can lead to momentum effects.

Similarly, if investors are overconfident, they may trade more and cause there to be a link between trading volume and market returns.

And finally, if investors are impacted by behavioral factors, then changing investor sentiment could impact market prices.

Trends and Corrections

Much of technical analysis seeks to uncover trends in market prices, which is essentially a search for momentum (absolute or relative). Three methods used to identify trends include:

- Dow Theory — A classic example of this is the so-called Dow Theory, which suggests that stock prices are driven by primary trends (the long term movements in prices), secondary trends (intermediate term movements away from the long term trends, which are eliminated via corrections) and minor trends (the daily fluctuation of little importance).

  Variations of the Dow Theory such as the Elliott Wave Theory and Kondrateiff Waves are similar.

- Moving Averages — Prices relative to a long term moving average may predict a movement towards its true value (moving average) or may identify changes in long term trends. The moving average essentially smoothes out short-term fluctuations and may make it easier to spot changes in the trends.

- Breadth — This is measured by comparing the number of winners, those with advancing prices, and losers, those with declining prices. It indicates the extent to which movement in the index is reflected widely in all of the stocks in the market.

Sentiment Indicators

Some examples of tools used to capture changing investor sentiment include:

- Trin Statistic — Technical analysts use trading volume to capture the degree to which market movements are significant — whether there is broad investor participation. The Trin Statistic uses the average volume of declining stocks (volume of declining stocks over the number of declining stocks) as a ratio to the average volume of advancing stocks. Ratios above one are considered bearish (negative) because they are thought to indicate net selling pressure.

- Confidence Index — This uses the bond market to gauge investor sentiment by calculating the ratio of high grade corporate bond yields to intermediate grade corporate bond yields. When investors are more optimistic, default premiums will shrink and the ratio will move closer to one.

- Put/Call Ratio — The ratio of outstanding put options to outstanding call options, which hovers around 0.65, is thought to be a signal of market sentiment. An increase in the ratio is thought to be bearish as it indicates more investors might be using put options to hedge their risk of a falling market. But some investors want to be buying when others are pessimistic, so they would view an increase in the ratio as a buying opportunity.

Practice Questions

Question 1. What is the behavioral critique of conventional financial theory?

Solution. The premise of behavioral finance is that conventional financial theory ignores how people really make decisions under uncertainty. First, there is experimental evidence that suggests that people commonly make errors inferring probability distributions. Second, even when they know the probability distributions, they often make systematically suboptimal decisions.

These are interesting observations, but their impact on security markets is less clear. If these sorts of errors lead to mispriced securities, then we should expect arbitrageurs to take advantage of these errors. Their actions would eventually cause the errors to be corrected. So if these errors do impact market prices, there must be some other limits to the actions of the arbitrageurs.

Question 2. The text categorizes various behavioral issues and discusses how these could impact market efficiency. List the three main categories discussed and briefly describe their implications.

Solution. The three major categories are:

1. *Information processing errors* make it difficult for investors to accurately assess probabilities and evaluate risk.

2. *Behavioral biases* make it difficult for investors to rationally evaluate risky opportunities using a simple rule to maximize their expected utility. Instead, they fall back on simple rules of thumb or are influenced by more subtle factors, such as regret avoidance.

3. *Limits to arbitrage* make it difficult for investors to exploit the mistakes made by others, allowing market prices to deviate from their fundamental value for long periods of time.

Question 3. Briefly describe four examples of information processing errors.

Solution. The four discussed in the textbook are:

1. Forecasting Errors — People tend to give too much weight to recent evidence and tend to produce extreme forecasts. This could lead to excessive earnings forecasts and cause high P/E stocks to subsequently under-perform.

2. Overconfidence — People tend to exhibit extreme overconfidence and overestimate their abilities.

3. Conservatism — Investors are often slow to update their prior beliefs in the face of new information.

4. Sample Size Neglect and Representativeness — People tend to infer patterns from limited information.

Question 4. Briefly describe four examples of behavioral biases.

Solution. The four discussed in the textbook are:

1. Framing — Decisions tend to be influenced by the way they are framed. For instance, people tend to be risk averse when it comes to gains (would prefer a certain gain to a gamble) but are risk seeking when it comes to losses (would prefer a gamble to a certain loss).

2. Mental Accounting — A specific type of framing is when people segregate certain decisions. For example, some gamblers may make different decisions when playing with the house's money versus when they have to dip into their own funds.

3. Regret Avoidance — People tend to blame themselves less when their choices are more conventional, which causes people to shy away from out of favor stocks. This could explain why higher returns are required to entice investors to buy firms with high book to market ratios.

4. Prospect Theory — This theory argues that investors often exhibit loss aversion. Their utility does not depend on the level of wealth, but rather on changes in wealth. This means that people do not get less risk averse as their wealth rises and they are more risk-seeking rather than risk averse when it comes to losses.

Question 5. Briefly describe three limits to arbitrage.

Solution. The three discussed in the textbook are:

1. Fundamental Risk — All strategies to exploit an apparent arbitrage opportunity likely contain some residual risk, such as the time it might take for the mispricing to be corrected in the market.

2. Implementation Costs — Often times, mispricings cannot be fully exploited due to transaction costs or other limitations.

3. Model Risk — What if you are wrong? All opportunities to exploit identified mispricings will carry some risk and therefore may limit the willingness of arbitrageurs to try to exploit it fully.

Question 6. List three real-world instances where there appeared to be violations of the Law of One Price. Comment briefly on how these violations might be explained.

Solution. The three discussed in the text are:

1. Siamese Twin Companies — Royal Dutch and Shell's share prices were linked by a simple formula that split the combined profits 60%/40% to the two firms, suggesting that Royal Dutch's price should always be 1.5 times Shell's price. Yet over long periods of time this was actually not the case, with deviations ranging from more than -5% to 17%.

   The fundamental risk that the mispricing could get even more out of whack made it difficult at times for these price differentials to be exploited.

2. Equity Carve Outs — In 2002 3Com sold 5% of their Palm subsidiary to new investors and arranged for existing 3Com shareholders to receive the other 95% of Palm's shares

(specifically, they were to receive 1.5 shares of Palm for each share of 3Com they owned). As a result, 3Com's price should have been at least 1.5 times the price of Palm prior to the distribution of the Palm shares to 3Com's shareholders, plus the value of all of 3Com's other businesses. However, for an extended period of time, 3Com's stock price was actually less than the price of Palm.

Apparently investors were overvaluing Palm and/or undervaluing 3Com, making it possible for an arbitrage profit by purchasing 3Com and shorting Palm. However, it was actually quite difficult and expensive to short Palm at the time, due to the very small volume of Palm shares outstanding and the fact that they were already sold short.

3. Closed-End Funds — The market value of closed-end mutual fund shares often reflects a substantial discount from the net asset value (NAV), which suggests that if the fund were to simply sell all of its holdings, the investors would receive a windfall since the holdings would be sold for the NAV.

   Researchers have found that the discounts on different funds tend to move together and are correlated with small stock returns, suggesting that all might be impacted by variations in investor sentiment. In addition, while it may be theoretically plausible to buy the funds trading at a discount and sell (short) the funds trading at a premium, eventually earning an arbitrage profit when prices converge to NAV, this would carry significant fundamental risk since the premiums and discounts could widen. And finally, the premiums and discounts could simply reflect a rational analysis of the dividends, expenses and alphas. Even minor amounts of expenses in excess of the fund's alpha can lead to substantial discounts.

BKM Chapter 13: Empirical Evidence on Security Returns

This chapter examines whether the CAPM and APT models developed earlier provide accurate estimates of expected returns. This discussion highlights many subtle aspects of these models. Pay particular attention to the Fama & French analysis.

Actual vs. Expected Rates of Return

Keep in mind that the CAPM and the APT model are both models of the **expected** return on a stock (as a function of the expected returns on other factors). Because expected returns cannot be observed (nor, by the way, can correlations or the exact market portfolio), all tests of these models necessarily require the use of actual returns from prior periods. This means that all tests of these models are already somewhat flawed.

This distinction is very important and is related to the distinction we made earlier between Index Models and CAPM. Remember, a single factor index model looks a lot like CAPM, it says that:

$$R_i = a_i + b_i R_M + \epsilon$$

But this is not a model explaining the expected return. It is a model explaining the relationship between the actual return for the stock and the actual return for the overall market. Therefore, tests of CAPM and APT wind up using index models. That's acceptable, in general, because the beta in the index model (in the special case where the index is the market return) will be measured the same way as the beta in CAPM.

Test Procedure

Suppose we wanted to test CAPM. How would we do it? The book begins with an explanation of a specific two-stage regression method that first calculates the betas for various stocks or portfolios of stocks and then uses those (estimated) betas to test the relationship between the actual historical return and the (estimated) beta.

Test Methodology Using Two-Stage Regression

The following steps are used in the two-stage regression:

1. *Setting Up the Sample Data* – First, gather historical return data for say, the past 60 months. Use a broad market index like the S&P 500 to represent the market and gather historical monthly returns for 100 different stocks.

2. *Estimating the Security Characteristic Line* – Then do the *first pass* regression using a single index model with the S&P return as the index. The point of this step is to get estimates of beta for each stock, so 100 different regressions will be performed with 60 data points in each regression.

The result will be a data set with 100 beta estimates, 100 estimates of the average excess return over the risk free rate for each stock, 100 estimates of the variance of the residuals and one estimate of the average excess return on the market.

3. *Estimating the Security Market Line* – Now, do a *second pass* regression using the average excess returns, betas and residual variances for the 100 stocks. In this regression, you will have one data point for each stock reflecting the average excess return during the 60 month time period for that stock (denoted in the text as $\overline{r_i - r_f}$), an estimate of the beta, an estimate of the residual variance (reflecting all of the non-systematic risk) and an estimate of the excess market return for the 60 month time period.

   The regression will be of the form:

   $$\overline{r_i - r_f} = \gamma_0 + \gamma_1 \hat{\beta}_i + \gamma_2 \hat{\sigma}_{ei}^2 + \epsilon$$

4. *Interpreting the Results* – The point of the regression just described is to test three things:

   a. test if the intercept equals zero,

   b. test if the slope equals the average excess market return ($\overline{r_M - r_f}$), and

   c. test if the coefficient on the residual variance term is zero (indicating that beta is the only source of risk that impacts average excess returns).

   If these tests are positive, then CAPM is proven to hold.

Results of Empirical Tests of CAPM (Lintner and Miller & Scholes)

When this analysis was performed, the results were somewhat consistent with CAPM. However, researchers found that $\gamma_0 > 0$, meaning that the intercept of the model was too high. Also, γ_1 was too small, making the slope of the line too flat.

This was disappointing, though not entirely surprising given the volatility of asset returns and therefore the lack of precision of statistical tests of average returns. Further, there are several issues that make the tests themselves are somewhat flawed:

- the tests don't use the "true" market portfolio
- investors cannot borrow or lend at the risk free rate, which is assumed in the simple versions of the CAPM
- asset volatility causes large measurement errors in the betas in the first pass regression

We will explore each of these issues in more detail below.

The Market Index

An important observation about the early tests of CAPM came from Richard Roll. In what has become known as Roll's Critique, he argued that the CAPM cannot be tested because the proxies used for the market portfolio do not come close to the portfolio of all risky assets called for by the model. Tests of the CAPM are really just tests of whether the market proxy is mean-variance efficient (whether it lies on the efficient frontier) or not.

Tests that reject a positive relationship between average return and beta may simply point to the fact that the market proxy used in the test is not on the efficient frontier. CAPM could be valid and tests using highly diversified, but not mean-variance efficient, portfolios could still fail the tests outlined above.

Kandel and Stambaugh showed that this critique can be extended to tests of the Zero-Beta CAPM as well. The parameters of the second-pass regression are biased by an amount that depends on how close the "market" portfolio is to the efficient frontier.

So where does this leave us? We cannot ever truly test CAPM because we do not know the "true" market portfolio and cannot reliably use proxies for it. We can, however, test the APT more readily though, because that doesn't rely on a market portfolio; the APT only relies upon the use of a well-diversified index portfolio. So we can still test the Security Market Line (SML) in an APT context (subject though to the measurement errors issues discussed below).

More on Roll's Critique

The key point of Roll's criticism of early CAPM tests was noted above, but it's worthwhile to summarize his entire argument here:

- The only test that can be done of CAPM is whether the theoretical market portfolio is on the efficient frontier.

- If the market portfolio is efficient, then there will be a linear relationship between expected return and beta – by definition. So tests of whether the relationship is linear are really just tests of whether the theoretical market portfolio is on the efficient frontier.

  Interestingly, the problems associated with the tests of CAPM noted above may not matter. If the linear relationship between expected return and beta follows directly from whether the market portfolio is efficient, then we need only show that this is the case. Since it is very hard for professional investors to beat the S&P 500 or the NYSE index, perhaps this alone is sufficient evidence for the empirical content of CAPM.

- After the fact, there are an infinite number of efficient portfolios, even if the theoretical market portfolio is not efficient. If any of these efficient portfolios are used as the market portfolio, then the betas estimated from this will result in a linear security market line and thus make it appear as though CAPM holds.

- CAPM can never be tested unless the true market portfolio (containing all risky assets) is used in the test.

- The proxy portfolio could be efficient even though the theoretical market portfolio is not, or vice versa. Since there is a high degree of correlation among the various choices of valid proxy portfolios, it may seem like the choice doesn't matter. In fact, the choice does matter because different proxy portfolios (different benchmark portfolios) may both result in a linear SML but could have wildly different implications for portfolio selection.

Measurement Error in Beta

When the right-hand side variables in a regression are measured with error (as the first-pass betas are in the tests above), the resulting parameter estimates for the slope are biased downward and the estimates for the intercept are biased upwards. This is consistent with the results mentioned above.

Miller & Scholes confirmed that the inability to accurately measure beta in the first pass regressions makes these tests deficient. They simulated return data that explicitly behaved like the CAPM predicts and yet the tests still failed.

Black, Jensen and Scholes (BJS) found that you could fix some of the statistical errors in the beta estimates, at least in theory. To test the model and minimize these types of errors, you need to use portfolios rather than individual stocks. But this has the downside of minimizing the number of data points you can use, since it isn't possible to form a large number of independent portfolios. Since the small number of data points will itself introduce statistical error, it is important to ensure that the small number of portfolios used have the greatest possible dispersion in beta estimates.

To clarify, suppose we had 100 stocks in our sample. Rather than randomly assigning these to five "well-diversified" portfolios of 20 stocks each, we rank each asset by its estimated beta and form portfolios with the 20 highest beta assets, the next 20 highest betas, etc. This gives us 5 portfolios with small non-systematic components and widely spaced betas, which should improve the ability to test the Security Market Line (SML) statistically.

Fama & MacBeth applied this procedure to test: i) if there was a relationship between average excess returns and beta, ii) if that relationship was linear (they included both a β term and a β^2 term to test the linearity of the relationship) and iii) if non-systematic risk impacts average returns. Their initial analysis using return data from 1935-1968 found that the relationship between return and beta is indeed linear (the coefficient on the β^2 term was not statistically significant) and that non-systematic risk does not explain excess returns (the coefficient on the residual standard deviation from the first pass regression was large but not statistically significant – it had lower t-statistics).

Note that the Fama& MacBeth results were reasonably favorable to the CAPM SML, but when data since 1968 is added the results are worse. In addition, Fama & MacBeth used an equally-weighted stock index in their initial analysis. If their analysis is updated using a value-weighted index, the results are worse even for the time period they originally tested.

The Final Verdict on CAPM Tests

The early CAPM tests confirmed the linear relationship between expected return and beta and confirmed that non-systematic risk does not affect expected return (i.e. non-systematic risk is not rewarded). However, the actual models are not fully verified because the intercepts of the tested models are too high and the slopes are too flat.

Tests of the Multifactor CAPM & APT

Tests of the single-factor SML perform poorly, but this may be because we have omitted relevant factors, besides just covariance with market returns, that could cause investors to demand risk premiums. Some examples might be:

- factors that hedge consumption against uncertainty in prices (e.g. we may demand higher risk premiums if asset returns are lower at a time when consumption costs (housing costs, energy costs or general inflation) are higher)
- factors that hedge future investment opportunities (e.g. interest rates or the market risk premium)
- factors that hedge assets missing from the market index (e.g. labor income or private businesses)

We will start with an extension of the SML to include labor income and private businesses, two sources of cash flows that are not reflected in the market index.

Labor Income

The standard CAPM assumes all assets (all sources of cash flow that investors can ultimately use for consumption of goods and services) are tradable. However, the income investors receive from their employment, referred to as their *labor income* or the value of their *human capital*, is not tradable and not reflected in the market index.

Notice that in the aggregate labor income is likely to be highly correlated with the market index (stock prices are likely to rise and labor income is likely to increase when companies are performing well). This will mean that estimated betas measured against an index comprised only of tradable asset are likely to be overstated. When a second-pass regression is performed against these overstated betas, the slope of the SML is likely to be too flat. This is what most of the SML tests discussed above had observed.

Suppose investors are worried about their labor income and want to hedge their risk by buying portfolios that perform well when aggregate labor income (as a proxy for their own labor income) declines. These investors are willing to pay more for such a portfolio that hedges their labor income risk, accepting a lower return on portfolios that are negatively correlated with aggregate labor income. Similarly, they will pay less for portfolios whose cash flows are low at the same time that labor income declines, demanding a higher rate of return on portfolios that are positively correlated with aggregate labor income. Taken together, this results in an SML with an additional risk factor.

Jaganathan and Wang (JW) did an empirical analysis of such a model using changes in aggregate labor income as a proxy for human capital. In their model, they also took into account the fact that betas are likely not fixed and might fluctuate with the business cycle (the state of the economy), using the spread between low and high grade corporate bond yields as a proxy for the business cycle. They also included a term to reflect the size of the company (or portfolio of companies).

They then ran a series of tests with different subsets of their selected factors. When they included only the standard market index beta, their second-pass regression coefficient on the index beta was insignificantly different from zero and the R^2 was very low. By adding the size of the firm(s) to the standard CAPM model, they improved the R^2 but the coefficient on the CAPM beta remained insignificant. Replacing the size factor with the labor income and business cycle factor produces a similar result, with a comparable R^2 and still an insignificant CAPM beta. And finally, adding in the size factor with the labor income and business cycle factors results in the size factor being insignificant.

The important implications of the JW study were:

- Standard betas do not adequately capture the cyclicality of stock returns and thus do not fully capture the systematic risk.
- Human capital is an important component of any model that tries to capture systematic risk.
- Test of whether firm size was a significant factor showed that it was not once the other JW variables were added to the standard CAPM.

Accounting for Non-Traded Assets

Recall that CAPM assumes that only systematic market risk impacts the expected returns for individual stocks. However, to the extent that investors hold significant investments in private firms that are not publicly traded, we should expect this to impact their portfolio choices (for the same reasons that we expect their human capital to impact these choices as well). Stocks whose returns have a high covariance with non-traded business income should have higher risk premiums.

Heaton and Lucas studied this issue and found that investors with more non-traded (private) business income tended to be less heavily invested in the overall market.

They also added changes in non-traded business wealth as a variable in the JW model described above and found that this variable was statistically significant. They also had the same result as JW in that the market index return did not help to explain the return on individual stocks, rejecting the core implications of the CAPM.

Multifactor CAPM & APT – The Chen, Roll and Ross Study

The important implication of the multifactor CAPM or the APT is that expected asset returns will be higher when they are positively correlated with priced risk factors. Unfortunately, there is no theoretical basis to specify the risk factors, or even to specify how many risk factors there may be.

Chen, Roll and Ross (CRR) performed an early empirical analysis wherein they identified several possible variables, in addition to the market index, that might proxy for systematic factors:

- growth in industrial production
- changes in expected inflation
- unexpected inflation
- unexpected changes in risk premiums as measured by differences in returns for corporate and government bonds
- unexpected changes in the term premium as measured by the difference between long-term and short-term government bond returns

CRR followed a similar approach as was described earlier in the Fama & Macbeth study. They created 20 portfolios by market value of outstanding equity and first estimated the betas for each of the 20 portfolios using 5 years of monthly data. They then used the 20 sets of betas as the independent variables in a second-pass regression and estimated the risk premiums on each of these betas. The market index wasn't statistically significant (and when a value-weighted index was used the sign of the risk premium was actually negative). In fact, the only three variables that were significant were industrial production, the risk premium on corporate bonds and unanticipated inflation.

Fama-French 3-Factor Model

Fama and French developed a multifactor model that includes the market, a firm size factor and a book-to-market ratio factor:

$$E(r_i) - r_f = b_i[E(r_M) - r_f] + s_i E(\text{SMB}) + h_i E(\text{HML})$$

Risk Factors vs. Firm Characteristics

The Fama-French model was motivated by the observation that small firms, as measured by their market capitalization, tended to have higher alphas relative to the CAPM than did larger firms. One may be inclined to believe that this is because simply having the *characteristic* of being small somehow results in higher returns. But it may also just be that small firms are highly correlated with a priced risk factor. For instance, smaller firms may be more sensitive to changes in future investment opportunities, such that an economic downturn systematically effects small firms differently than large firms. In this sense, the alphas observed relative to CAPM may simply be compensation for a priced risk factor, a risk premium, for firms that are more sensitive to economic downturns.

To capture this risk factor, and others as well, Fama-French did not just add a term to reflect the size of the firm and its book to market ratio. Instead, they created indices that could serve as proxies for the relevant (but unknown) risk factors. For any particular portfolio, what matters is how sensitive the returns are to those two indices.

The Fama-French Indices

The size index, denoted SMB, reflects the difference in returns for small firms minus the returns for big firms. It is constructed by sorting all firms by their market capitalization and grouping those into two portfolios representing Small firms and Big firms. The groupings are done at the beginning of each year, the monthly returns for the Small and Big portfolios are calculated and the difference is recorded as the monthly SMB size premium.

The book to market index, denoted HML, is constructed similarly using firms ranked by their book-to-market ratios and put into three groups, High, Medium and Low. (Note that firms with high ratios of book value to market value are referred to as ”value” firms and firms with low ratios of book value to market value are referred to as ”growth” firms.) The rankings are done annually and then the monthly returns for the High and Low portfolios are used to calculate the monthly book to market premium (also referred to as the value premium).

These calculations are clarified in Table 1.

TABLE 1. Fama-French Portfolios

	Size		
Book-to-Market Ratio	*Small*	*Big*	Average
High	Portfolio S/H	Portfolio B/H	**A**
Medium	Portfolio S/M	Portfolio B/M	B
Low	Portfolio S/L	Portfolio B/L	**C**
Average	**D**	**E**	

The return for the Small size portfolio is the average of the returns for the three portfolios in the first column (i.e. for all B/M ratios). This average is depicted in the table as **D.** The return for the Big size portfolio is the average of the returns for the three portfolios in the second column, depicted as **E** in the table. The SMB value for the particular period is then the difference of these two average returns, **D - E**. Similarly, the HML value for the particular period is the difference between the averages in the high row and the averages in the low row, or **A - C** in the table.

Testing the FF Model

Using the SMB and HML indices, along with the monthly excess returns for a very broad market index, the following three-factor model is tested using a time series regression of actual returns for each test portfolio against the excess market return, SMB and HML indices:

$$r_i - r_f = a_i + b_i[r_M - r_f] + s_i\text{SMB} + h_i\text{HML} + \epsilon_i$$

In this regression equation, the i subscripts indicate which portfolio is being used, not the time period. For example, in one study by Davis, Fama and French (DFF) they studied **nine** portfolios formed using three size groups (small, medium and large) and three book-to-market groups (high, medium and low). They were then able to produce one set of a, b, s and h parameters for each of the nine portfolios (recall that the SMB and HML indices were created separately using six portfolios and were not impacted by how many portfolios were tested).

Figure 1 shows the parameter estimates and their t-statistics for the nine portfolios in the DFF analysis using data from 1929 to 1997:

FIGURE 1. Fama-French Results

	B/M	Size	Excess Return	*a*	*b*	*s*	*h*	*t*(*a*)	*t*(*b*)	*t*(*s*)	*t*(*h*)	R^2
S/L	0.55	22.39	0.61	−0.42	1.06	1.39	0.09	−4.34	30.78	19.23	1.73	0.91
S/M	1.11	22.15	1.05	−0.01	0.97	1.16	0.37	−0.18	53.55	19.49	9.96	0.96
S/H	2.83	19.05	1.24	−0.03	1.03	1.12	0.77	−0.73	67.32	39.21	26.97	0.98
M/L	0.53	55.85	0.70	−0.06	1.04	0.59	−0.12	−1.29	55.83	18.01	−4.30	0.96
M/M	1.07	55.06	0.95	−0.01	1.05	0.47	0.34	−0.15	32.98	17.50	9.50	0.96
M/H	2.18	53.21	1.13	−0.04	1.08	0.53	0.73	−0.90	47.85	8.99	11.12	0.97
B/L	0.43	94.65	0.58	0.02	1.02	−0.10	−0.23	0.88	148.09	−6.88	−13.52	0.98
B/M	1.04	92.06	0.72	−0.09	1.01	−0.14	0.34	−1.76	61.61	−4.96	13.66	0.95
B/H	1.87	89.53	1.00	−0.09	1.06	−0.07	0.84	−1.40	52.12	−0.86	21.02	0.93

TABLE 13.6

Three-factor regressions for portfolios formed from sorts on size and book-to-market ratio (B/M)

Source: James L. Davis, Eugene F. Fama, and Kenneth R. French, "Characteristics, Covariances, and Average Returns, 1929 to 1997," *Journal of Finance* 55, no. 1 (2000), pp. 396. Reprinted by the permission of the publisher, Blackwell Publishing, Inc.

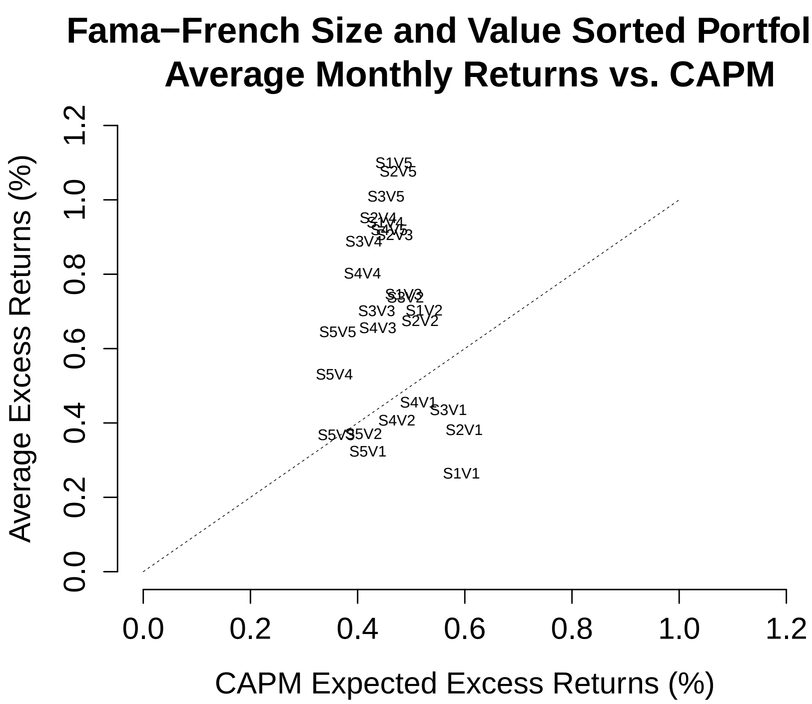
Based on the results shown in Figure 1, the models have a large R-squared and the t-statistics for the factor loadings (b, s, h) show that these factors contribute to the explanatory power of the model. Notice that within a size bucket (the first three rows), the h loadings and the average returns increase as you move from the L, M to H value portfolios. This is true in each of the size buckets (S, M, B). Similarly, the s loadings and average returns decrease as you move from S to M to B size groupings (though not as cleanly and less in the later time periods than in the earlier time periods).

Fama-French 3-Factor Model vs. CAPM

Two graphs similar to Figure 13.1 in the textbook should make it clear how the Fama-French 3-Factor Model compares to CAPM in terms of being able to explain historical returns.

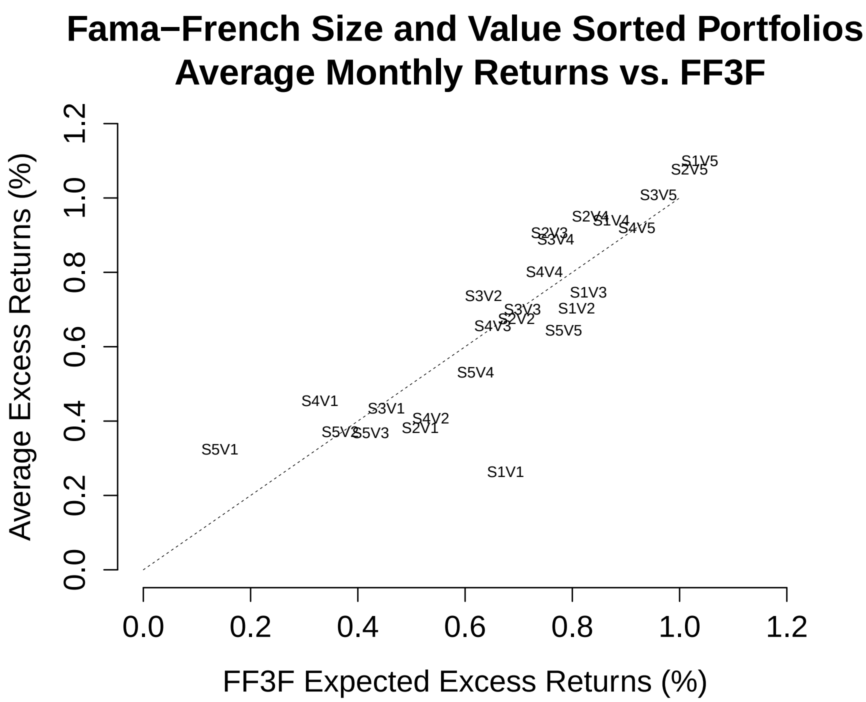
In the first chart, I use data from 1947-2013 (through August) to estimate the CAPM beta for 25 different portfolios. The portfolios have been selected by sorting all U.S. stocks according to their size (five groups, labeled S1 to S5, with 1 being the smallest) and their book-to-market ratios (five groups, labeled V1 to V5, with 1 being the lowest book-to-market ratios). For instance, the portfolio with the highest book-to-market ratios among the smallest firms would be labeled "S1V5". The sorts are redone each year and then the return for the subsequent year is calculated.

I then compare the CAPM's expected excess return to the actual average excess returns for the 25 portfolios. If the CAPM is an accurate model, we would expect all portfolios to lie along the 45-degree line in the chart.



As you can see, the CAPM performed pretty poorly, overstating returns for the lowest book-to-market firms and understating them pretty significantly for the largest book-to-market firms.

As an alternative, the same analysis is done using the FF3F model, which performs far better:



What Does it Mean?

The results summarized above seem to suggest that the CAPM is inadequate and that the size and book-to-market ratios are important determinants of average returns. While that may be true, it is difficult to understand why this is the case. Are the SMB and HML indices serving as proxies for a risk not fully captured in CAPM, or are the risk premiums the result of investor irrationality and behavioral biases? It turns out that we don't really know for sure.

There are compelling arguments on both sides of this debate, which fall into two categories – risk-based interpretations and behavioral interpretations – which we will address below.

Risk Based Interpretations

One argument is that the size and value factors must be priced factors for risks not captured in CAPM. Portfolios with a high degree of sensitivity to these factors earn higher returns on average, which is what we'd expect if investors demanded higher expected returns for taking on risk. Only here, risk is defined more broadly than just covariance with the market portfolio.

There are two bits of evidence that the SMB and HML factors are proxies for a source of risk not captured in CAPM:

- There is evidence that the SMB and HML indices capture the risk in the business cycle. The returns on the HML and SMB indices (the risk premiums) seem to be positively correlated with future growth in the macroeconomy. Investors evaluating particular stocks would demand a higher risk premium if returns were going to be bad when the overall economy was bad, so the correlation between the SMB and HML indices and the overall economy explains at least part of the risk premiums that are observed.

- We often assume that betas and market risk premiums are constant. But what if they varied over time, or more importantly, if the betas were high when the risk premiums were high? For instance, high book to market firms tend to have more tangible capital (factories, equipment, etc.) and therefore they may be more susceptible to the effects of recessions. Their large fixed investments compared to firms with less tangible capital can make them less responsive to economic downturns because they cannot easily reverse their prior investment plans (hiring less people or firing some of them is much easier than selling specialized equipment). As a result, during times of recession the betas of these high book to market firms might be high. Since there is also evidence that during recessions the market risk premium is also higher, the effect could be that the betas and the risk premiums are correlated. If this were the case, then we would expect to see greater overall risk premiums for high book to market firms.

To test the effect of positive correlation between the market risk premium and the betas of high book-to-market firms, researchers assumed that both the betas and the market risk premiums were driven by common *state variables* that can capture the state of the economy such as the market dividend yield, the default spread on corporate bonds, the slope of the term structure and the risk free rate. This allowed them to model different betas and market risk premiums in each time period and to use the estimated market risk premiums as a proxy for the overall economy.

Using that model, they find that the betas of high book to market firms (so-called *value stocks*) were higher during recessions and lower during expansions than the betas of low book to market firms (so-called *growth stocks*). This was even more significant for the top 10% and bottom 10% observations of the market risk premium (corresponding to troughs and peaks, respectively, in the economy).

This explains why the HML index may be serving as a proxy for an underlying source of risk. When using the Fama-French model, we are not necessarily concerned with modeling the stock's specific book to market ratio – we're really just modeling the sensitivity of the stock price to this HML index. That's our way of modeling the stock's sensitivity to the underlying risk that the HML index is a proxy for.

Behavioral Interpretations

Another way to interpret the evidence that the HML portfolios help to explain returns on stocks, thus resulting in a *value premium*, is that investors may irrationally prefer certain glamour firms that have had good recent performance, high prices and thus lower book-to-market ratios. This causes their returns to lag behind the value firms.

The behavioral rationale for this might be due to the representativeness error discussed in Chapter 12, which causes investors to extrapolate recent performance too far into the future and to overreact to good news. The evidence that this is the case is that when firms are ranked by their past growth rates in income, their book-to-market ratios decline steadily. This suggests that investors extrapolate high growth too far into the future, overvalue the firm (see BKM Chapter 18 for valuation models and how growth impacts them) and cause the book-to-market ratios to fall. As a result, high book to market firms will tend to be those whose prices were not exaggerated and their returns will, on average, be higher.

Notice that if low book-to-market firms (the glamour firms) did indeed have higher growth, investors would be right to value them highly. However, when the low book-to-market firms are identified up-front, their future five year growth rates are not higher than other firms. That is, it does seem as though the book-to-market ratios reflect past rather than future growth.

Additional evidence comes from an analysis of earnings announcements for firms classified as either growth (low book-to-market ratio) or value stocks (high book-to-market ratio). On average, investors are relatively disappointed with the earnings announcements of growth stocks. It seems as though they optimistically bid up the price of these stocks (causing their book-to-market ratios to be low) based on expectations of good earnings and were generally disappointed when actual earnings were announced.

Momentum: A Fourth Factor

Recall in BKM Chapter 11 we discussed the tendency for good or bad performance (measured on a relative basis) of stocks to persist, producing a *momentum* effect. A momentum factor (WML, or Winners minus Losers) is therefore often added to the FF Three Factor model when evaluating past performance, say of mutual funds.

There is, however, a significant challenge interpreting this momentum factor. While researchers have had modest success justifying the SMB and HML factors as proxies for priced risk factors, it's more difficult to do this with the momentum factor.

Liquidity and Asset Pricing

The liquidity extension of CAPM was introduced in Chapter 9. That model showed that expected returns are impacted by liquidity in two ways – through consideration of the effect of transaction costs (bid-ask spreads, price impact, etc.) and through consideration of the liquidity risk that results from covariance of the changes in liquidity costs with changes in market liquidity costs and market risk premiums.

Liquidity itself is hard to measure, but it is possible to construct an index of market-wide illiquidity and then calculate liquidity betas. When stocks are then sorted into ten portfolios reflecting the liquidity beta deciles, we find that both CAPM and FF three-factor alphas are higher for the high liquidity beta portfolios and low for the low liquidity beta portfolios. This supports the notion that liquidity effects indeed play a role in explaining returns for stocks. One study that included momentum in a four-factor FF model had similar results, and even suggested that much of the momentum effect could in fact be driven by liquidity risk.

And recall the analysis in BKM Chapter 9 that used multiple liquidity betas within the standard CAPM to improve our ability to explain returns.

Equity Premium Puzzle

In a 1985 article, Mehra and Prescott found that over the period 1889-1978, investors have been excessively rewarded for bearing risk. The excess market returns observed in the U.S. seemed too large to be consistent with economic theory and reasonable levels of risk aversion, which suggests that we should be very careful about using historical excess market returns as forecasts of future risk premiums.

There have been several attempts to address this puzzle:

- Mehra and Prescott used a Consumption CAPM (CCAPM), but their analysis suffered from data issues that have since been improved upon, mitigating some of the perceived puzzle.
- Investors just got lucky – they weren't expecting (or demanding) high market risk premiums, but they were fortunate to earn them. In this sense, there really is no puzzle.
- The excess returns may have been real, but the measure is biased because it only reflects U.S. market data that suffers from a survivorship bias.
- Extensions to the standard CAPM which remove some of its less realistic assumptions also resolves some of the puzzle.
- Liquidity risk considerations may also mitigate some of the perceived puzzle.
- Behavioral biases could lead to irrational behavior, which in turn causes the equity premium puzzle.

We'll explore each of these arguments below.

Consumption Growth and Market Returns

Recall that CAPM assumes that the only thing that matters to investors is their total wealth. Since basic portfolio theory shows that all investors will hold the market portfolio of risky assets (perhaps levered up or down with borrowing or lending), they evaluate risk for specific stocks in terms of the covariance of their returns with the returns on the overall market portfolio.

The ICAPM model introduced in Chapter 9 argued that what really matters to investors is not their wealth *per se* but their lifetime consumption (what their wealth can buy). The market risk premium should be a function of investors' risk aversion, A, and the covariance of market returns with aggregate consumption. If market returns and consumption growth are not highly correlated and aggregate risk aversion is not too high, then the market risk premium should be much lower than what we have realized historically.

Testing this, via the Consumption CAPM (CCAPM) is difficult:

- it has been hard to capture consumption data directly (though researchers have had better luck using consumption-tracking portfolios),
- consumption changes tend to be concentrated in the fourth quarter of the year, and
- each consumer/investor is likely to have different consumption habits.

Recent research has addressed these issues and produced interesting results based on the CCAPM.

One study showed that the Fama-French factors are correlated with consumption betas and that the average returns for the 25 Fama-French portfolios (the 25 portfolios formed using quintiles of stocks sorted by size and quintiles of stocks sorted by book-to-market ratios) are strongly associated with their consumption betas.

Another study distinguished between different classes of investors and found that the covariance of market returns and consumption is far higher when we focus only on the consumption risk of households that actually own financial securities.

This work suggests that the historical market risk premium may not represent as much of a puzzle as was originally thought.

Expected vs. Realized Returns

Fama and French (in a different paper than the one discussed above regarding their 3-Factor Model) looked at the risk free rates, S&P 500 returns and the resulting equity risk premium realized from 1872–1999 and saw that the equity premium increased dramatically in the 1950–1999 period, suggesting that the puzzle is really a recent issue.

Noting the important distinction between the actual returns observed and the expected returns measured in various security pricing models, they attempted to measure the expected returns using a dividend discount model (see BKM Chapter 18). According to that model, the price of a stock is equal to the expected future dividends discounted at the investors' expected return:

$$P = \frac{Div_1}{k - g}$$

where k is the expected return and g is the growth rate in dividends. Rearranging, we can write the expected return as $k = Div_1/P + g$. In other words, the expected return is equal to the dividend rate plus the growth rate in dividends, as follows:

$$\text{Expected Return in Period } t = \frac{Div_t}{P_{t-1}} + g$$

In contrast, the *actual* return is equal to the following:

$$\begin{aligned}\text{Actual Return in Period } t &= \frac{Div_t + P_t}{P_{t-1}} - 1 \\ &= \frac{Div_t}{P_{t-1}} + \frac{P_t}{P_{t-1}} - 1\end{aligned}$$

In other words, the actual return is equal to the dividend yield plus the realized capital gain. And therefore the difference between actual returns and expected returns is reflected in the difference between the expected dividend growth rate (g) and the realized capital gains ($P_t/P_{t-1} - 1$).

What Fama and French noticed was that prior to 1950, the expected and realized (actual) returns were about the same, but that after 1950 the realized returns far exceed their estimated expected returns. The puzzle may simply be the result of larger than expected capital gains during this period.

It is interesting to ask which of these methods produce more reliable predictions of the future – the historical actual returns or the dividend discount model's (DDM) expected returns? Fama and French conclude it is the latter, for the following reasons:

- realized returns couldn't plausibly be representative of what firms expected without implying that they were knowingly engaging in negative NPV projects;
- estimates from the DDM are measured much more accurately; and
- DDM estimates are much more consistent with stable investor risk aversion.

Their conclusion from this analysis is that the observed rates of return were due to unexpected capital gains during the latter part of the period studied (1950–1999) that cannot be expected to continue in the future. In addition, when returns are extended back to 1792, we see further evidence that the 1950–1999 period really was unusual and that the equity premium puzzle as usually documented was driven by actual realized returns during a rather short period of time.

Survivorship Bias

Because the equity puzzle was based on the returns for the U.S. stock market, the results are likely to be biased. The U.S. market has been unique in that it has been by far the most successful, has had the highest return, has been in existence the longest and has never had serious problems of being shut down for extended periods (or closed as has happened in some countries). Judging the returns on risky assets based solely on this exceptional market is misleading.

Extensions to Standard CAPM

Another way to reconcile the observed market risk premiums being higher than we would expect using the standard CAPM, ICAPM or the CCAPM is to recognize that investors face significant risks associated with loss of income. They therefore may rationally demand higher risk premiums due to risks associated with the business cycle. In addition, young people cannot really borrow without constraints, which impacts the demand for bonds and the resulting equity risk premium. And investors may become more risk averse as their consumption declines (referred to as "habit formation"). Incorporating this feature into the assumed utility functions also explains part of the puzzle.

Liquidity and the Equity Premium Puzzle

We saw earlier that liquidity risk is important in explaining historical stock returns. Part of the average excess return we observe is almost certainly compensation for liquidity risk rather than just systematic return variability.

Behavioral Explanations

Narrow framing is a commonly observed behavioral trait in which people tend to consider risky decisions in isolation instead of as part of an aggregate portfolio and therefore don't give adequate consideration to the fact that their stock portfolio risk may have low correlation with the rest of their wealth. Combined with loss aversion, which causes investors to exaggerate the pain of losses, this can lead to large equity risk premiums. That is, investors *effective* risk aversion is higher than traditional measures of risk aversion suggest.

Practice Questions

Question 1. The following table summarizes various data collected over the past 60 months for 200 stocks as well as for the S&P 500 Index and the US T-Bill rate.

TABLE 2. Historical Monthly Returns

Month	T-Bill	S&P 500	Stock 1	Stock 2	Stock 3	⋯	Stock 200
1	0.29%	1.64%	1.31%	0.87%	0.96%	⋯	1.23%
2	0.39%	1.20%	1.11%	0.68%	0.85%	⋯	0.84%
3	0.22%	1.46%	0.88%	0.83%	0.84%	⋯	1.25%
4	0.18%	1.28%	0.79%	0.93%	1.49%	⋯	1.08%
5	0.38%	1.01%	1.06%	1.37%	0.64%	⋯	0.91%
⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮
57	0.33%	0.81%	0.76%	1.30%	0.90%	⋯	0.82%
58	0.30%	1.11%	1.26%	0.69%	1.33%	⋯	0.71%
59	0.29%	1.17%	1.03%	1.02%	1.33%	⋯	0.85%
60	0.26%	1.03%	0.72%	0.82%	0.59%	⋯	0.91%
Average	0.28%	1.18%	1.00%	0.92%	0.82%	⋯	0.96%

Assume that you are attempting to test the CAPM by running a two-pass regression. Answer the following questions: a) what is the model for the *first pass regression*; b) how many observations will there be in your first pass regression; c) what is the first data point you will use in this first pass; d) what is the model for the *second pass regression*; e) how many observations will there be in your second pass regression; f) what is the first data point you will use in this second pass?

Solution. The first pass regression is used to get estimates of beta for each stock. It uses a market model of the form:

$$r_i - r_f = a + b(r_{Mi} - r_f) + e$$

where r_i is the return for the stock in month i and r_{Mi} is the return for the market in month i (using the S&P 500 index as a proxy for the market). This first pass will have a total of 60 data points, one for each month, and will be done separately for all 200 stocks. The first data point will use the 1.31% return for Stock 1, 1.64% return for the market and .29% for the risk free rate r_f. The first data point will then be $1.02\% = a + b(1.35\%) + e$.

The second pass will then use these 200 beta estimates in another model of the form:

$$\overline{r_i - r_f} = \gamma_0 + \gamma_1 \hat{\beta}_i + \gamma_2 \hat{\sigma^2_{ei}} + \epsilon$$

The purpose of this pass is to get parameters for a model relating the average returns on each stock in excess of the risk free rate as a function of its beta. There will be 200 data points, one for each stock's return and beta. The first data point will use 1.00% – .28% = .972% for the dependent variable and the beta estimate from the first pass for Stock 1.

Question 2. Assume that γ_0 in the model above is statistically different from zero and positive. Is this consistent with the Standard CAPM? What form of CAPM might it be consistent with?

Solution. The standard CAPM predicts that this term would equal zero. However, under the assumption that investors cannot borrow and lend at the risk free rate, then the Zero Beta version of CAPM is probably better. This result would be consistent with this zero beta version of the model.

Question 3. Assume that in the second pass regression discussed above γ_2 was non-zero. Is this consistent with the standard CAPM model? If not, what important aspect of CAPM does it violate? What alternative model might be more consistent with these results?

Solution. If this parameter were non-zero, then it would not be consistent with CAPM. It suggests that non-systematic risk does in fact affect security returns while CAPM assumes that only systematic risk, beta, matters.

However, it does not necessarily mean that the residual variance itself matters. The residual variance may just be serving as a proxy for other relevant risk factors, which would make the results consistent with a multi-factor CAPM or APT model.

Question 4. BKM point out that early empirical tests of CAPM are generally inconsistent with CAPM. List two of these inconsistencies.

Solution. The intercept of the tested models is too high (above the risk free rate); the slope is too flat (lower than the average excess market return).

Question 5. One of the concerns about interpreting any tests of CAPM is that it represents a model of expected returns and we can only observe actual returns on highly volatile stocks. This makes the precision of these tests quite low. Identify three other fundamental concerns about these tests.

Solution. The tests necessarily have to use a proxy for the theoretical market portfolio; the statistical accuracy of the beta estimates is suspect and there is substantial sampling error; the assumption about risk free borrowing and lending is obviously unrealistic.

Question 6. Despite the findings of the empirical research that raises questions about certain aspects of the CAPM, what two major conclusions of CAPM are largely supported?

Solution. Expected returns increase linearly with systematic risk, β; expected returns are not affected by non-systematic risk.

Question 7. Fama and French's three-factor model suggests that the total risk premium earned on a portfolio or individual stock consists of a market risk premium as in CAPM but also a size and value risk premium related to their SMB and HML factors. In addition, they found that portfolios with high factor loadings on the HML factor (high value for h_i) earned higher average returns. Since the HML factor is based on the difference in returns between high book to market and low book to market portfolios, it would seem that a firm with a high book to market ratio should earn a higher expected risk premium. Explain why this is not necessarily the case.

Solution. The FF three-factor model does not say that firms with high book to market ratios will earn higher expected risk premiums. The HML factor is used as a proxy for some combination of underlying risk factors. Firms whose returns are sensitive to or correlated with this factor have a higher factor loading (h) and it is this high factor loading that results in higher returns on average. It is not the case that high or low book to market firms necessarily have either high or low factor loadings.

To better understand this, consider the table shown in Figure 2 taken from the Davis, Fama and French paper referenced in the textbook.

FIGURE 2. Relationship Between Average Returns and HML Factor Loadings

	BE/ME	Size	Ex Ret	a	b	s	h	$t(a)$	$t(b)$	$t(s)$	$t(h)$	R^2
Low BE/ME												
S/L/Lh	0.51	22.29	0.49	−0.56	1.21	1.25	−0.02	−2.88	16.53	6.93	−0.22	0.70
S/L/Mh	0.57	22.79	0.69	−0.34	1.07	1.21	0.15	−2.15	16.26	12.47	1.27	0.77
S/L/Hh	0.56	21.05	0.76	−0.38	1.03	1.64	0.26	−1.91	10.96	7.21	2.03	0.73

Recall that Davis, Fama and French created 9 portfolios, one of which contained the stocks that had Small size and Low book-to-market ratios (denoted the S/L portfolio). We can further break this particular portfolio into three portfolios with low, medium and high values for the h loadings. Notice that the expected returns seem to rise as the h factors rise, despite the fact that the book-to-market ratios and the firm sizes are roughly the same across these three portfolios. Therefore, what drives the differences in returns is the differences in the h loadings and not the size or book-to-market characteristics of the stocks in the portfolios.

Question 8. What are the two possible interpretations of the Fama and French Three-Factor Model results?

Solution. One argument is that the size and value factors must be priced factors for risks not captured in CAPM. For instance, they may capture risks associated with the business cycle or the risks associated with time varying betas and market risk premiums. Portfolios with a high degree of sensitivity (correlation) to these factors earn higher returns.

Another way to interpret the same results is that investors irrationally prefer firms with good performance, high prices and lower book-to-market ratios, perhaps because they incorrectly extrapolate good recent performance too far into the future.

Question 9. Suppose that liquidity risk was a priced risk factor, meaning that in equilibrium liquidity risk does indeed impact expected returns. Is it possible that the SMB and HML risk premiums in the Fama-French 3-Factor model are affected by liquidity risk?

Solution. The SMB and HML risk premiums were merely observed factors, formed through a simple portfolio construction rule, which Fama and French have *interpreted* to serve as proxies for risk.

If indeed liquidity risk were a priced factor, then perhaps the reason for the large historical SMB risk premium is that small stocks might be less liquid than big stocks, resulting a risk premium for those who are willing to own the less-liquid stocks.

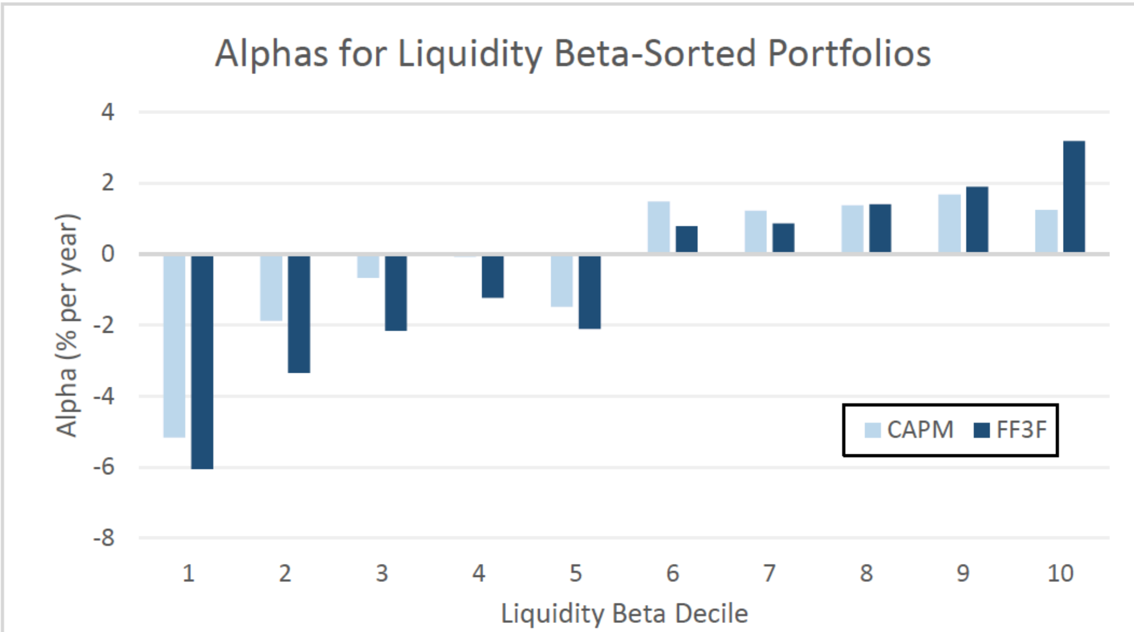
Similarly, if liquidity risk were a priced risk factor, then perhaps the reason for the large historical HML risk premium is that high book to market firms might be more susceptible than low book to market firms to recessions or other indicators of “bad times". During those times, a lack of liquidity can be particularly painful for investors, resulting in an implicit liquidity risk premium in the HML index.

Note that the 2014 exam had a question (Question 6) related to liquidity risk and the Fama-French model. In the Examiner’s Report, they specifically noted that a response like the one I just gave was “incorrect” and did not receive credit. Obviously, I think the graders got this one wrong.

Question 10. Suppose, as in the previous question, that liquidity risk is a priced risk factor but that the SMB and HML indices do not capture this liquidity risk. If you tested the Fama-French model using data for a sample of firms, what would you expect to see?

Solution. If indeed liquidity risk were not implicitly captured in the SMB and HML indices, then a test of the Fama-French model would produce non-zero alphas reflecting the average liquidity risk premium. For example, in the textbook (see Exhibit 13.6), they show the CAPM and FF alphas for firms sorted based on separately measured liquidity betas:

FIGURE 3. Fama-French Alphas and Liquidty Betas



Note that Question 6 on the 2014 exam essentially asked about this topic. However, the wording of the question "Describe the impact of liquidity on the Fama-French three-factor model" was vague. First, BKM state explicitly that the FF three-factor model ignores liquidity. Second, if liquidity were not a priced factor then there would be no effect on the FF3F model. Third, as noted in the previous question, a valid response should have been that liquidity risk might be among the many possible explanations for the historical SMB and HML risk premiums. And finally, the effect they intended you to talk about was on the FF3F model alphas, and not the model itself.

Question 11. Describe the equity premium puzzle and briefly discuss the four explanations for this phenomenon given in the text.

Solution. It appears as though the realized returns in excess of the risk free rate over the past 100 years have been much higher than any rational security pricing model would suggest.

However, Fama and French pointed out that this is really a recent phenomenon over just the past 50 years and that a crude measure of expected returns using the dividend discount model is actually much lower than the actual returns, suggesting that investors have just been lucky.

Secondly, the puzzle is based only on the returns in the US stock market rather than on all global markets including those that have since shut down. The inclusion of only the most successful market creates a survivorship bias.

Third, when CAPM's assumptions are revised to reflect borrowing constraints for young consumers, the fact that consumers face risks associated with their labor income and the inclusion of habit formation into utility functions, we can get higher risk premiums than CAPM suggests.

And finally, loss aversion and narrow framing can cause investors to have higher effective levels of risk aversion that leads to higher market risk premiums than traditional measures of risk aversion would suggest.

Recommended Textbook Problems

I strongly recommend working all of the end-of-chapter questions from the BKM textbook. But due to time constraints this may not be feasible. At a minimum you should review all of the Concept Checks.

Selected Old Exam Questions for Part 2

The following questions relevant for this section appeared on the Old CAS Exam 8 from 2000 to 2010 and on the CAS Exam 9 since 2011.

BKM 9	BKM 10	BKM 11	BKM 12	BKM 13
2001 Q10	2000 Q13	2003 Q15	2007 Q7	2000 Q12
2002 Q11	2001 Q32	2004 Q10	2008 Q6	2001 Q11
2003 Q10	2002 Q15	2006 Q8	2013 Q5	2003 Q16
2003 Q11	2003 Q13	2007 Q4		2004 Q11
2004 Q4	2004 Q9	2007 Q5		2005 Q10
2004 Q5	2006 Q7	2008 Q5		2005 Q11
2004 Q6	2008 Q4	2008 Q7		2006 Q9
2004 Q7	2009 Q3	2012 Q3		2014 Q6
2005 Q12	2011 Q2	2014 Q2		2015 Q6
2005 Q7	2015 Q4b			2015 Q7
2005 Q9				
2006 Q6				
2008 Q3				
2010 Q4				
2011 Q3				
2011 Q4				
2013 Q4				
2014 Q3				
2015 Q4a				
2015 Q5				

For some of these questions I have provided the text of the question and an explanatory solution. These were selected either because they are representative of the questions you are likely to be asked on future exams or because they contain an element that is particularly worthwhile to point out. For the other questions, the CAS solutions should be sufficient to confirm whether your answer is correct.

Important Note: The solutions shown here are intentionally detailed. They contain thorough explanations of the concepts and formulas used to reinforce the main points from the readings and provide an additional teaching opportunity. **Actual exam responses should be much more concise than what is shown here, along the lines of what you will see in the solutions that the CAS releases.**

2007 Exam Question 5

a. Briefly describe the semi-strong version of the Efficient Market Hypothesis (EMH).

The Semi-Strong Form of the Efficient Market Hypothesis states that all publicly available information regarding the prospects for the firm is reflected in the stock price. It is therefore not possible to earn abnormal (risk-adjusted) returns simply by using publicly available information.

b. Identify and briefly describe three anomalies that appear to contradict the semi-strong version of the EMH.

If markets are efficient, then we should not be able to find simple trading rules based on publicly available information that lead to excess risk adjusted returns. However, researchers have found a number of examples of so-called anomalies that seem to defy the EMH.

- P/E Effect — Evidence suggests that low P/E stocks outperform high P/E stocks, even after adjusting for differences in betas between the various groups. This would be contrary to the EMH because ranking stocks based on their P/E ratio is so easy that we would expect people to do this without much effort and then force up the price of low P/E stocks and force down the price of high P/E stocks. But in reality, perhaps beta is simply not fully reflecting the risk differences in the low P/E and high P/E stocks.

- Small Firm in January Effect — Banz showed that ranking firms by their total market capitalization we can see that small firms tend to significantly outperform large firms. Again, even after adjusting for differences in betas, this small firm effect remains and is rather significant. This too would contradict the EMH since such a simple rule should not be able to result in such excess profits.

  It turns out that virtually all of the small firm effect occurs in January, and usually in the first few days. One explanation of this is the impact of tax loss selling which tends to cause mostly small stocks to be sold in December to lock in tax write-offs and then are repurchased in early January.

- Neglected Firm Effect — Researchers found that firms that are followed less by analysts (e.g. smaller firms) tend to have less readily available information and tend to earn excess risk adjusted returns. This is likely because the lack of readily available information makes these firms riskier in ways that the CAPM beta itself does not capture.

  Recall from Chapter 9 that illiquidity effects can cause the prices of smaller, neglected firms to be significantly reduced, resulting in a rate of return premium. This could partially explain both the small firm in January effect and the neglected firm effect. However, it wouldn't explain why the small firm returns should be high only in January. Nonetheless, the high trading costs of illiquid small and neglected firms could wipe out any apparent opportunities to earn abnormal returns.

- Book to Market Ratios — Fama and French have found that firms with the highest ratios of book value to total market capitalization tend to outperform firms with the lowest ratios. More importantly, they found that once both the size and the book to market effects are taken into account, the standard CAPM beta seems to have no power to explain average returns.

- Post Earnings Announcement Price Drift — Often there is a sluggish response to earnings announcements. In one study, analyst estimates of earnings were compared to actual earnings to estimate the earnings surprise and then 10 portfolios were formed with the smallest to largest surprises. The abnormal returns for each portfolio were then compared for the 60 days before and after the earnings announcement.

  For firms with positive surprises, the cumulative abnormal returns are positive on the announcement day. For firms with negative surprises, the CAR is negative. However, the CARs tend to drift further upwards or further downwards after the announcement date. This suggests that people don't fully reflect the surprises fast enough and that a simple strategy of buying the positive surprise firms and selling the negative surprise firms would have been profitable.

2007 Exam Question 7

Briefly describe three information processing errors that can lead investors to make mistakes estimating probability distributions in the financial markets. In each case, provide an example of a market anomaly that could be explained by these errors.

Note that I have slightly reworded the question to avoid some unintended confusion.

There are four main information processing errors identified by behavioral finance researchers that can all lead to errors estimating probability distributions.

- Forecasting Errors — People tend to give too much weight to recent evidence and tend to produce extreme forecasts. This could lead to excessive earnings forecasts and cause high P/E stocks to subsequently under-perform.

- Overconfidence — People tend to exhibit extreme overconfidence and overestimate their abilities. This can lead to a greater reliance on active portfolio management, as opposed to passive portfolio management, than proponents of the EMH would predict and to excessive trading.

- Conservatism — Investors are often slow to update their prior beliefs in the face of new information. This can lead to momentum or serial correlation in stock prices.

- Sample Size Neglect and Representativeness — People tend to infer patterns from limited information. This causes them to overreact and then correct their mistakes only gradually, causing observed reversals in stock price changes.

2006 Exam Question 6

The following table gives a security analyst's opinion about expected returns on two stocks for two different market scenarios:

Scenario	Market Return	Stock A	Stock B
1	6%	2%	5%
2	22%	34%	11%

The risk free rate is 4% under both scenarios.

Note — The sample CAS solution is not entirely correct. This question is essentially identical to Question 9 in BKM (Tenth Edition) Chapter 9. Refer there for a slightly different version of the question that may be less confusing.

a. Calculate the beta for Stock A and Stock B.

In Part (a), they didn't want you to assume that CAPM was valid and then solve for the beta. Instead, they wanted you to take the information given by the analyst and estimate the beta as the sensitivity of the return to changes in the market return. In other words, they wanted the slope of the regression coefficient between the market return and the stock returns (if there is a linear relationship between these points, what is the slope of that relationship?).

This is the way that beta is actually calculated, but usually there are more than two points and usually it is based on historical actual returns rather than an analyst's estimates.

With more data points (e.g. if we were using historical data) then this would be solved using standard regression calculations. With the type of data given here though, you simply need to connect the two points on the graph with the market return as the "X" variable and the stock return as the "Y" variable.

Plot the two points (6%, 2%) and (22%, 34%) and then just calculate the slope as the rise $(34\% - 2\%)$ over the run $(22\% - 6\%)$ to get beta of $32/16 = 2.0$ for Stock A and beta of $6/16 = 0.375$ for Stock B.

b. Calculate the alpha for each Stock A and Stock B if Scenario 1 and Scenario 2 are equally likely.

If the scenarios are equally likely, the expected returns are 14% for the market, 18% for Stock A and 8% for Stock B.

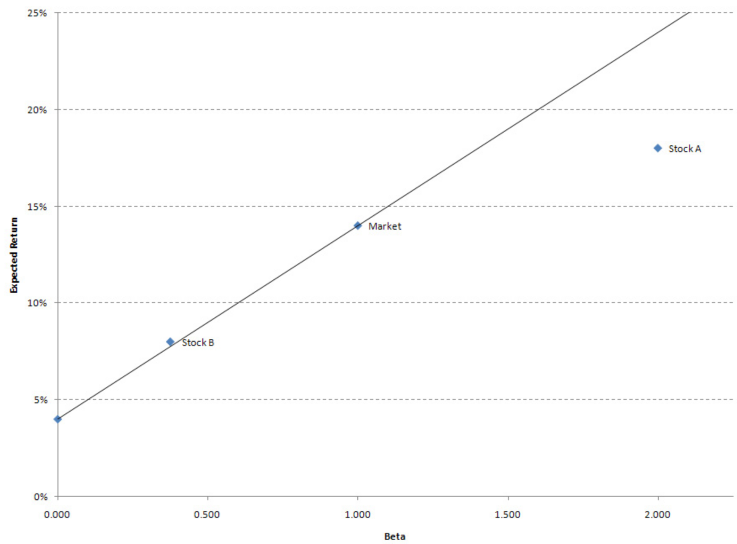
Further, if the betas estimated from Part A are correct and CAPM holds, then CAPM would say that the expected return for Stock A should be $E(r_A) = 4\% + 2(14\% - 4\%) = 24\%$. Since the expected return is 18%, this is an alpha (deviation from the CAPM line) of -6%.

For Stock B, the alpha is found similarly to be .25%.

c. Draw the Security market Line (SML) for the given economy and plot the two securities on the SML graph. Label all items properly.

For the SML, just plot the points using the risk free rate as one point, with a beta of zero, and the expected market return and the market beta of 1 as the other point. Then the two securities will be graphed OFF the line, since the expected returns are below/above the SML.

FIGURE 1. Comparison of Stocks A and B to SML



Note that the CAS sample solution appears to mistakenly plot all of the points on the line.

d. Briefly explain which stock, A or B, would be perceived by the analyst as a better buy.

The analyst would view Stock B as better because it plots just above the SML and therefore has an expected return that is higher than its beta suggests it should be.

However, note again that this is entirely dependent on the analyst estimating beta using his own estimates of the expected returns, so this is somewhat strange.

2004 Exam Question 9

Assume there are two independent economic factors, T and S. All stocks have independent firm-specific components to their returns with a standard deviation of 40%.

You have the following information on three well-diversified portfolios:

	Portfolio 1	Portfolio 2	Portfolio 3
Beta T	2.0	0.8	1.2
Beta S	-0.5	1.0	1.0
Expected Return	13%	10%	???

The risk free rate is 4%.

Calculate the expected return of Portfolio 3.

This is an APT question identical to one in the notes. You are told that there are two factors and are given the returns and betas for three portfolios (Portfolio 1, Portfolio 2 and the risk free portfolio). These three portfolios can be used to derive the three factor portfolios (the risk-free portfolio and the portfolios with beta of 1.0 to one factor and zero to the other) and then the factor risk premiums are simple to estimate. These are then plugged in to estimate the return for Portfolio 3.

The end result of an APT model is going to be of the form:

$$E(R_P) = r_f + \beta_{iT}[E(r_T) - r_f] + \beta_{iS}[E(r_S) - r_f]$$

We already know $r_f = 4\%$, so we really just have two equations and two unknowns. These are:

$$13\% = 4\% + 2[E(r_T) - 4\%] + (-.5)[E(r_S) - 4\%]$$

$$10\% = 4\% + .8[E(r_T) - 4\%] + 1[E(r_S) - 4\%]$$

These are easily solved for $E(r_T) = 9\%$ and $E(r_S) = 6\%$.

Then plug these back into the first equation with Portfolio 3's betas to get:

$$E(r) = 4\% + 1.2[9\% - 4\%] + 1[6\% - 4\%] = 12\%$$

2002 Exam Question 15

Suppose that asset returns are given by a two-factor model and four portfolios exist with the following expected returns and factor betas:

Portfolio	Expected Return	Beta 1	Beta 2
A	15%	0.5	1.2
B	12%	0.4	0.8
C	11%	0.2	1.0
D	4%	0.0	0.0

a. Using this information, determine the Arbitrage Pricing Theory (APT) formula for the expected returns on an efficient portfolio.

The first part of this requires you to find the parameters of the APT model. You are told that there are two factors and therefore you know that any three diversified portfolios can be used to solve for the returns on the factor portfolios.

Recall the resulting APT equation is:

$$E(R_i) = \lambda_0 + \lambda_1 b_{i1} + \lambda_2 b_{i2}$$

and so we simply have to write this equation for any three of the given portfolios and then simultaneously solve this fairly complex algebra problem for the three unknown lambda values.

It's unfortunate that you had to do the algebra on the actual exam, but it turns out that they made this a bit easy for you because they gave you Portfolio D with zero sensitivity to factor 1 or factor 2, so we know that $\lambda_0 = 4\%$.

Now we just have two equations and two unknowns. You can solve them to get $\lambda_1 = 10\%$ and $\lambda_2 = 5\%$ using any of the two remaining portfolios. You could have also used any other combination of three portfolios, but not using Portfolio D would have taken you more time.

b. What will happen if there is a portfolio E with an expected return of 12% and factor betas of .45 and .90 respectively for factors 1 and 2?

The second part asked what would "happen" if there was another portfolio E with $E(R) =$ 12%. In this case, the APT model suggests that it should have an expected return of $4\% + .45(10\%) + .9(5\%) = 13\%$. Therefore, an arbitrage opportunity would exist and someone would take advantage of it. Eventually this opportunity would go away.

That's the literal answer to a vague question, but the full answer is that the arbitrage will involve SELLING portfolio of E (since it has poor returns relative to other combinations of portfolios with the same factor sensitivities) and BUYING the combination, call it G, with the same sensitivities.

If we did this say with \$100 worth of each portfolio, the combination portfolio G would have an expected payoff of \$113 and the short position in E would have an expected payoff of -\$112. The net payoff would be \$1, with no risk.

This would be done until the pressure to sell E caused its price to fall and its returns to rise. When everything settles down, the arbitrage should be gone.

2008 Exam Question 4

You are given the following information:

- There are three independent factors F_1, F_2 and F_3
- The risk free rate is 3%
- The risk premium for F_3 is 10%
- You are given the following information regarding three well diversified portfolios:

Portfolio	Beta on F_1	Beta on F_2	Beta on F_3	Expected Return
A	0.80	0.7	1.65	20.00%
B	-0.25	1.8	0.30	11.90%
C	1.25	0.7	1.40	16.60%

Determine the risk premium for F_1.

This question is essentially the same as the previous question (2002 Exam Question 15) with just a minor twist to make it seem harder but in practice actually makes it easier.

Recall the resulting APT equation when there are three factors is given by the following formula:

$$E(R_i) = \lambda_0 + \lambda_1 b_1 + \lambda_2 b_2 + \lambda_3 b_3$$

Since we know the risk premium for the third factor is $\lambda_3 = 10\%$ and that the risk free rate λ_0 is 3%, all three of the equations for the expected return on portfolios A, B and C can be written in terms of just two variables, λ_1 and λ_2. Then, any two of these equations could be used to solve for the unknown risk premium λ_1.

It turns out that in this particular case, the algebra is trivial if you select portfolios A and C, since they have the same beta on F_2 and so they can be readily combined to eliminate one of the two variables.

After doing this trivial algebra, it turns out that the risk premium on $F_1 = -2\%$.

2008 Exam Question 6

a. Define the Law of One Price

When two different assets with identical cash flows and identical risks can be bought or sold simultaneously at two different prices, an arbitrage opportunity exists. Someone could buy the asset at the low price and simultaneously sell an economically identical asset at the higher price, netting the difference as a risk free profit. Since such opportunities, if found, would quickly be exploited, the Law of One Price states that two identical assets with identical risks should not be available at different prices.

b. Give two examples that violate the Law of One Price

The three discussed in the text are:

- Siamese Twin Companies — Royal Dutch and Shell's share prices were linked by a simple formula that split the combined profits 60%/40% to the two firms, suggesting that Royal Dutch's price should always be 1.5 times Shell's price. Yet over long periods of time this was actually not the case, with deviations ranging from more than -5% to 17%. The fundamental risk that the mispricing could get even more out of whack made it difficult at times for these price differentials to be exploited.

- Equity Carve Outs — In 2002 3Com sold 5% of their Palm subsidiary to new investors and arranged for existing 3Com shareholders to receive the other 95% of Palm's shares (specifically, they were to receive 1.5 shares of Palm for each share of 3Com they owned). As a result, 3Com's price should have been at least 1.5 times the price of Palm prior to the distribution of the Palm shares to 3Com's shareholders, plus the value of all of 3Com's other businesses. However, for an extended period of time, 3Com's stock price was actually less than the price of Palm.

  Apparently investors were overvaluing 3Com, making it possible for an arbitrage profit by purchasing 3Com and shorting Palm. However, it was actually quite difficult and expensive to short Palm at the time, due to the very small volume of Palm shares outstanding and the fact that they were already sold short.

- Closed-End Funds — The market value of closed-end mutual fund shares often reflects a substantial discount from the net asset value (NAV), which suggests that if the fund were to simply sell all of its holdings, the investors would receive a windfall since the holdings would be sold for the NAV.

2015 Exam Question 4b

Given the following information:

- The variance-covariance matrix:

	Stock A	Stock B	Market Portfolio
Stock A	0.1600		
Stock B	-0.4800	1.4400	
Market Portfolio	-0.0108	0.2520	0.0900

- Stock A and Stock B are perfectly negatively correlated.
- The risk free rate is 3%
- The expected return on the market portfolio is 8%.

Construct an arbitrage strategy using the risk-free rate and some combination of Stock A and Stock B.

Because A and B are perfectly negatively correlated, it is possible to construct a portfolio of those two stocks that has no risk (zero variance). The weights can be found by first noting that because of the negative correlation, the variance of the risky portfolio can be written as follows:

$$\sigma_P^2 = (w_A\sigma_A - (1 - w_A)\sigma_B)^2$$

Just set that equal to zero and solve for $w_A = 75\%$ and $w_B = 25\%$.

To see if there is an arbitrage, we need to calculate the expected return for this portfolio. Using the information given, we can calculate the betas for each stock and their expected returns. Then, using those and the weights above, we can determine the expected return on this portfolio is 6.05%.

Since that return is riskless and higher than the 3% return on the risk-free asset, we know that there is an arbitrage opportunity. We simply borrow at the risk-free rate to purchase the portfolio composed of 75% in Stock A and 25% in Stock B. At the end of the year, we will earn 6.05% and have to repay interest of 3%, leaving an arbitrage profit of 3.05%.

2015 Exam Question 6

Given the following information (in $ millions):

Firm	Book Value	Market Value	CAPM Beta
A	625	500	1.2
B	350	500	1.2
C	1250	1000	1.2

Use the principles underlying the Fama & French 3-Factor Model to evaluate which firm would be expected to have the highest future rate of return.

The empirical data on stock returns has shown, over a long history, for their to be a tendency for small firms, as measured by their market value, and for firms with high ratios of book value to market value to have higher average rates of return than large firms or firms with low ratios of book value to market value, even after adjusting for differences in CAPM betas.

The information given suggests that Firm A is the smallest firm with the highest ratio of book value to market value of the three firms shown. Therefore, we might expect, based on the historical data studied by Fama and French and no other information, that Firm A would have a higher average return in the future.

Keep in mind the point that I've made multiple times in the notes and in the practice questions that the above response has very little to do with the Fama-French 3-Factor Model. That model does **NOT** predict that small firms or high book to market firms will have higher expected returns. What it argues is that if you compile an index comprised of the difference in returns between small firms and large firms (SMB) and an index comprised of the difference in returns between high book to market and low book to market firms (HML), then portfolios that are positively correlated with those indices (have high loadings on the SMB and HML indices) will have higher expected returns than firms that are negatively correlated with those indices.

This is an important distinction between the empirical observation that Fama and French made and the 3-factor model that they developed to try to explain those observations.

Assume that the firm identified in Part A above does produce the highest rate of return over a future period. Describe two behavioral explanations that could explain the superior performance.

Fama and French have argued that the size and book to market premiums (excess returns) observed are risk-based and that their SMB and HML indices are proxies for risks not captured in the CAPM. For example, they might be capturing risks associated with the business cycle or the variations in betas and market risk premiums over time.

However, behavioralists argue instead that the excess returns are due to investor irrationality. As a result, there are behavioral explanations for the small firm and value effects.

With respect to the small firm effect, behavioralists would argue that small firms tend to be neglected by investors due to regret avoidance. These firms get less attention from the press and from stock analysts, causing investors to require a higher premium to make a less conventional investment.

With respect to the book-to-market effect, investors may irrationally extrapolate good results too far into the future and overreact to good news. This would cause them to prefer glamour firms with good recent performance and high prices, and thus low book-to-market ratios. If investors overpay for those firms, then we would expect their future returns to lag behind the returns for firms with high book-to-market ratios.

2015 Exam Question 7

An investor has held the same portfolio of stocks since 1955.

	Avg Risk-Free Rate	Avg Port Return	Avg Div Yield	Avg Div Growth Rate
1955-1974	4.10%	9.50%	6.60%	3.00%
1975-1994	3.80%	9.10%	4.90%	4.00%
1995-2014	4.30%	12.60%	5.60%	3.50%

Using the dividend discount model (DDM), evaluate whether the high returns in the most recent period can be expected to persist into the future.

When evaluating the equity premium puzzle, Fama and French pointed out that high realized returns could have been to "luck" and would not represent a puzzle unless investors **expected** the high returns that were observed. They then argued that the best way to infer what the expected returns were during a historical period would be to use the dividend discount model, which would produce measures of expected return equal to the dividend yield plus the dividend growth rate. When they did this, they saw that realized returns were much higher than the DDM estimates of expected returns, due for instance to changes in risk free rates and/or investor risk aversion. Since it is not realistic to assume those changes would occur again in the future, investors should expect returns more in line with the DDM than the historical average.

In this question, realized returns were close to the DDM estimates of expected return in the first two periods (9.5% vs. 9.6% from 1955-1974 and 9.1% vs. 8.9% from 1975-1994). In the last period, however, realized returns of 12.6% far exceeded the DDM estimate of 9.1%. We cannot tell from the facts in the question what triggered this, but we can assume that there was some unexpected capital gains. Since the risk-free rate was higher than in the prior period, it's hard to know what triggered these gains – but something likely occurred to change the valuation (decrease in required risk premiums, for instance). Since it is not realistic to assume future changes in valuations, it is more likely that future returns will mirror the DDM returns and be more like the historical average dividend yields and growth rates, suggesting that future returns will be lower than in the recent period.

Part 3

Asset-Liability Management

Part 3 Table of Contents

Background: Bond Pricing

Introduction

The current CAS 9 syllabus assumes students know how to value coupon bonds and are familiar with some basic terminology. Since this background is needed to do a lot of the numerical problems elsewhere on the syllabus, I am including this basic summary here for those who need it.

Overview of Bond Types

The term *bond* is used to describe a very wide variety of financial instruments in which one party (the issuer) borrows money from another party (the investor or the lender) and agrees to repay them over time. The bond is the contract which specifies the exact terms of the repayment.

There are several types of bonds that differ with respect to the terms of this repayment:

- Zero Coupon Bond — This is a bond that does not pay periodic interest and simply pays a fixed dollar amount at maturity. The amount paid at maturity is referred to as the *principal* on the bond or the *par value* of the bond.

- Coupon Bond — This is a bond that pays periodic interest payments, called *coupon* payments over the life of the bond and then the principal amount at maturity (along with the final coupon payment). Coupon payments can occur annually, semi-annually (every 6 months) or more frequently. Often the coupon rate, which indicates the percentage of the principal/par value that is paid each period, is stated as an annual rate such that a 10% coupon bond with semi-annual payments actually pays 5% of the principal value as interest every six months.

- Floating Rate Bond — A floating rate bond is a coupon bond that does not have a fixed coupon rate and instead pays different coupon amounts each period based upon a stated reference rate. For instance, a bond may pay coupons according to the current risk-free rate on each rate reset date. In most cases the rate will be reset at the *start* of each coupon period and then the payment at that rate will be made at the end of the period.

Each of these bond types appear on the syllabus and will be used throughout this manual.

Bond Pricing — Annually Compounded Interest Rates

Consider for simplicity a bond that pays annual coupons, C_i, and principal, P, at maturity. If we denote the interest rate appropriate for discounting a cash flow occurring at time time i as

r_i then the present value of the bond's stated cash flows is:

$$B = \sum_{i=1}^{n} \frac{C_i}{(1+r_i)^i} + \frac{P}{(1+r_n)^n}$$

Numerical Example

For example, if the one-year rate is $r_1 = 6\%$ and the two-year rate is $r_2 = 7\%$ then the price of a 5% coupon bond with annual coupon bonds and a par value of $\$1,000$ is:

$$\begin{aligned} B &= \sum_{i=1}^{n} \frac{C_i}{(1+r_i)^i} + \frac{P}{(1+r_n)^n} \\ &= \frac{\$50}{1.06^1} + \frac{\$50}{1.07^2} + \frac{\$1000}{1.07^2} \\ &= 964.28 \end{aligned}$$

Using a Financial Calculator

In the special case where the interest rate is the same for all maturities it is often much faster to calculate the price of a coupon bond using a financial calculator.

The bond pricing function assumes that there are n payment periods with a fixed cash flow each period equal to PMT as well as a final payment of the bond's face value equal to FV at maturity. When each cash flow is discounted at the same rate per period, denoted I/Y on the calculator, then the present value of the cash flows can be found by computing the PV by first pressing the CPT button and then the PV button.

The present value will be shown with the opposite sign as the PMT and FV amounts.

Numerical Example

For example, assume the interest rate for all periods is equal to 6% and that you wanted to value an 8% coupon bond with annual coupon payments, a face value of \$1,000 and 10 years until maturity. The following keystrokes can be used to get the price of this bond:

10 $\boxed{\text{N}}$ 6 $\boxed{\text{I/Y}}$ \$80 $\boxed{\text{PMT}}$ \$1000 $\boxed{\text{FV}}$ $\boxed{\text{CPT}}$ $\boxed{\text{PV}}$

This will return a value of \$1,147.20 (ignoring the negative sign).

Bond Yield

As a general rule, we will always allow for the interest rate, r_i, to vary each period and we will use the prices of zero coupon bonds maturing at time i to determine what this rate should be.

For convenience though, we will want to be able to refer to the *average* interest rate used to value the bond's cash flows. This average rate is the bond's *yield* and it is found be setting the bond price, calculated as shown above, equal to the present value of its *promised* cash flows when discounted at a constant rate, y:

$$B = \sum_{i=1}^{n} \frac{C_i}{(1+y)^i} + \frac{P}{(1+y)^n}$$

Numerical Example

Using the example referenced earlier with the 5% two-year coupon bond, we can set this equation equal to the price of \$964.28 and solve for $y = 6.97\%$. Obviously, for anything other than a zero coupon bond the algebra to solve for y will be a mess and so you will want to use your financial calculator. Just enter in all of the known information, including the PV as $-\$964.28$, and then solve for the yield by pressing $\boxed{\text{CPT}}$ $\boxed{\text{I/Y}}$.

As you can see, the yield is a weighted average of the 6% and 7% interest rates used to determine the price of the bond, with most of the weight on the two-year rate. Viewing the yield as a weighted average of the discount rates for each cash flow helps to make it clear that the yield on a zero coupon bond is just the interest rate for the cash flow that occurs at maturity, or $y = r_n$.

Determining the Price from the Yield

Although we will generally calculate the price from the zero-coupon bond yields and the promised cash flows and then determine the yield, in practice if you are given the promised cash flows and the yield you could also use that information to calculate the price.

In this case, as in the case where the interest rates were the same for all maturities, you can use your financial calculators to quickly determine the price of a coupon bond with multiple cash flows at regular time periods.

Coupon vs. Yield

An important point to note about the examples shown earlier is that there is no necessary relationship between the coupon rate for a bond and the interest rates or yields used to value the cash flows. The coupon rate is just a rate that is negotiated and agreed to when the bond is issued and then is fixed for the life of the bond. As interest rates fluctuate the coupon may be higher or lower than the rates used to calculate the present value.

In the special case where the coupon rate is equal to the yield, the bond will be worth its par value. However, if the coupon is lower than the yield then the bond will be worth less than its par value and will be said to be trading at a discount. Similarly, if the coupon is higher than the yield then it will be worth more than its par value and will trade at a premium.

Floating Rate Bonds

Because floating rate bonds, by definition, have coupon payments that are equal to the rate of interest required by investors (the coupon is set equal to the current interest rate), these bonds are always worth their par value *on the interest rate reset date*. Once the rate is set then the price can fluctuate slightly because the next coupon rate is now fixed, but upon the next interest rate reset date it will again be worth par.

Alternative Compounding Periods

In the previous discussion we assumed annual coupon bonds and annual interest rates. There are two other choices, besides annual, for the compounding frequency of the interest rate that we will encounter on the syllabus, semi-annual compounding and continuous compounding.

Semi-Annual Compounding

Because coupon bonds often pay their coupons semi-annually it is common for interest rates to be similarly quoted on a semi-annual basis. That is, a reference to an 8% interest rate, compounded semi-annually, means that the interest rate is 4% every six month period. An interest rate quoted in this fashion will often be referred to as being stated on a *bond equivalent basis*.

When the rate is quoted on a bond equivalent basis, the present value of a cash flow equal to \$1 in 6 months will be $\$1(1+r/2)^{-1}$ and a cash flow equal to \$1 in 12 months will be $\$1(1+r/2)^{-2}$.

In general, a cash flow at time n will be:

$$PV(CF_n) = \frac{CF_n}{(1+r/2)^{2n}}$$

Numerical Example

Assume you want to determine the value of a two year corporate bond with a face value of \$1,000, an annual coupon rate of 10% and coupons paid semi-annually. If the bond equivalent yield on similar bonds is 8%, then we can value the bond by calculating the present value of each of the cash flows.

The semi-annual coupon payments mean that the coupon is $10\%/2 = 5\%$ per 6-month period. Given the face value of \$1,000 the coupon payments every six months will be \$50. Also note that the interest rate to discount the cash flows is $8\%/2 = 4\%$ per 6-month period because the rate was given on a bond equivalent basis.

$$\begin{aligned} B &= \frac{5\%(1,000)}{1.04} + \frac{50}{1.04^2} + \frac{50}{1.04^3} + \frac{50+1,000}{1.04^4} \\ &= 1,036.30 \end{aligned}$$

This could have been found using your financial calculators with the following keystrokes:

4 $\boxed{\text{N}}$ 4 $\boxed{\text{I/Y}}$ \$50 $\boxed{\text{PMT}}$ \$1000 $\boxed{\text{FV}}$ $\boxed{\text{CPT}}$ $\boxed{\text{PV}}$

This will return a value of \$1,036.30 (ignoring the negative sign).

Continuous Compounding

The Hull readings generally assume continuous compounding whenever calculating present values (though bond-equivalent rates are usually used to reference the payment rates for swaps and forward rate agreements, as per market convention). With continuous compounding, the present value of a cash flow at time t is:

$$PV(CF_t) = CF_t e^{-rt}$$

Numerical Example

Assume you want to determine the value of a two year corporate bond with a face value of \$1,000, an annual coupon rate of 10% and coupons paid semi-annually. If the continuously compounded yield on similar bonds is 7%, then we can value the bond by calculating the present value of each of the cash flows as follows.

$$\begin{aligned} B &= 50e^{-.07(.5)} + 50e^{-.07(1.0)} + 50e^{-.07(1.5)} + 1050e^{-.07(2.0)} \\ &= 1,052.74 \end{aligned}$$

It would be nice to be able to use your financial calculators for problems such as this, but you need to be careful. The calculators assume a compounding frequency equal to the number of cash flows per period rather than continuous compounding. To use your calculators, you need to convert the continuous rate into a semi-annual rate as follows:

$$1 + r = e^{.07(.5)} \;\Rightarrow\; r = 3.562\%$$

Then, using this rate we can solve for the present value of the cash flows as follows:

4 $\boxed{\text{N}}$ 3.562 $\boxed{\text{I/Y}}$ \$50 $\boxed{\text{PMT}}$ \$1000 $\boxed{\text{FV}}$ $\boxed{\text{CPT}}$ $\boxed{\text{PV}}$

This will return a value of \$1,052.74 (ignoring the negative sign).

BKM Chapter 15: Term Structure of Interest Rates

The Yield Curve

When the yields on various bonds are plotted against the time to maturity for those bonds, the resulting diagram is called the yield curve. The yield curve will be an important input to our pricing of all bonds, but before we get to that it is important to distinguish between various types of yield curves.

The two most important yield curves for now are:

- The **pure yield curve** is perhaps the most important one for our purposes. It is created by compiling the yields on zero coupon Treasury bonds that pay a single cash flow at a single point in time. Treasury STRIPS can often be used to get these zero coupon bond yields. However, these are only available for selected maturities, so an interpolation procedure will also be needed to obtain the zero coupon yields for other maturities.

- The **on-the-run yield curve** is different in that it uses the yields on coupon bonds rather than zero coupon bonds and relies primarily on the most recently issued Treasury bonds (hence the term “on-the-run”).

The pure yield curve is particularly important because it can serve as the basis for valuing any stream of cash flows. The on-the-run yield curve is less useful because each of the bonds used contain a series of cash flows and so the overall yield is really a complex average yield for each of its cash flows. When valuing a specific cash flow, it is necessary to use the zero coupon, or pure, yield curve.

Then, to value any stream of cash flows, including a coupon bond, we simply break it into a series of separate cash flows that each occur at a specific point in time and use the zero coupon yields as the discount rate.

Short Rate vs. Spot Rate

Assume for a moment that we knew precisely what the 1-year interest rates were going to be one year from now, two years from now, etc. These various rates, called the *short rates*, represent the interest rate for a specified time interval (such as one year). These rates for different points in time (today, one year from now, two years from now, etc.) are denoted as r_1, r_2, r_3, etc., where the subscript can be interpreted as the maturity year for a bond that was issued one year prior.

The present value (PV) of \$1,000 to be paid in 1 year is found by simply discounting at the current 1 year rate, r_1. This would be written as:

$$PV = \frac{\$1{,}000}{1 + r_1}$$

The PV of $1,000 to be paid in two years would be discounted at the current one year rate and also at the one year short rate one year from now, or

$$PV = \frac{\$1{,}000}{(1 + r_1)(1 + r_2)}$$

We can use the prices we just determined for each of the zero coupon bonds with different maturities and solve for the yield, y_i. This yield would represent the *spot rate* for the relevant maturity. It represents the yield today on a zero coupon bond with the specified maturity. It is the collection of all of these spot rates at different maturities that gives us the pure yield curve described above (though technically we should be careful and refer to it as the *term structure* to avoid confusion with the on-the-run yield curve).

For example if $r_1 = 10\%$ and $r_2 = 11\%$, then the present value of a two-year zero coupon bond would be:

$$PV = \frac{\$1{,}000}{(1.10)(1.11)} = \$819$$

Using this, we can solve for the two-year spot rate by solving the following equation for y_2:

$$\$819 = \frac{\$1{,}000}{(1 + y_2)^2} \Rightarrow y_2 = 10.499\%$$

Spot vs. Forward Rates

Let's continue to assume that the future short rates are known with certainty and we wanted to invest $1,000 for three years. We could do this in two different ways:

1. Invest in a three year zero coupon bond at the current three year spot rate. This will give us a value of:

$$\$1{,}000(1 + y_3)^3$$

2. Invest in a two-year zero coupon bond at the two-year spot rate and then plan to reinvest the proceeds at the one-year rate two years from now (the future short rate). This will give us proceeds equal to $\$1000(1 + y_2)^2$ in two years which will then be reinvested at the future short rate to obtain a total payment of:

$$\$1{,}000(1 + y_2)^2(1 + r_3)$$

Since these two strategies should produce the same results (again, assuming no uncertainty with respect to the future short rates), then we can use the zero coupon spot rates y_i to solve for the future short rates r_i.

The relationship between them for this example is:

$$(1 + y_3)^3 = (1 + y_2)^2(1 + r_3)$$

Since the future short rates cannot really be known with certainty, to be careful we will refer to the future short rate calculated in this fashion as the *forward rate*, and rewrite this as:

$$(1+y_3)^3 = (1+y_2)^2(1+f_3)$$

or more generally,

$$(1+y_n)^n = (1+y_{n-1})^{n-1}(1+f_n)$$

Interest Rate Uncertainty and Forward Rates

So far we have assumed that we knew the future short rates and found that someone wanting to invest for two years could either invest for one year and roll it over for another year, resulting in $(1+r_1)(1+r_2)$, or they could just invest in a two-year zero coupon bond and get $(1+y_2)^2$.

If one of these strategies were better than the other, investors would prefer one over the other and the prices and yields would adjust. Therefore, it is fair to assume that these two strategies will ultimately produce the same results, or:

$$(1+r_1)(1+r_2) = (1+y_2)^2$$

But now we have to recognize that future rates are not certain, so we might want to rewrite this to say that the *expected value* of these strategies should be equal. Notice that the first strategy is actually uncertain because it depends on what the future short rate is but the zero coupon strategy can be locked in now. Therefore, we might want to say that:

$$(1+r_1)[1+E(r_2)] = (1+y_2)^2$$

Going further though and noting that only one of those strategies is risky, it may be the case that investors who want to invest for two years would still prefer the two-year zero coupon bond. Therefore, it may be necessary to entice two-year investors into buying a series of one year bonds by offering them higher returns, which would change this to an inequality:

$$(1+r_1)[1+E(r_2)] > (1+y_2)^2$$

Finally, note that we earlier defined the forward rate as the rate that would establish the following equality:

$$(1+r_1)(1+f_2) = (1+y_2)^2$$

and therefore if we plug this in to the previous equation we can see that in this case where we have to entice long term investors to buy short term bonds:

$$E(r_2) > f_2$$

Let's look at this from a different angle. Suppose instead we had an investor who wanted to invest $1,000 for just 1 year. This investor could do one of two things:

1. Invest in a one year zero coupon bond with a cash flow equal to:

$$\text{Cash Flow} = \$1{,}000(1 + r_1)$$

2. Invest in a two-year zero coupon bond with a cash flow in two years equal to $\$1{,}000(1+y_2)^2$ but then just sell it in one year. Its price at that time will depend on the one-year short rate at that time, which is currently unknown. The expected cash flow in one year is then:

$$\text{Expected Cash Flow} = \frac{\$1{,}000(1 + y_2)^2}{1 + E(r_2)}$$

As before we should expect these two different quantities to be equal, so:

$$(1 + r_1) = \frac{(1 + y_2)^2}{1 + E(r_2)}$$

For this investor, the one-year bond has no risk but buying a two-year bond and planning to sell it in one year has risk because of the unknown future short rate. In this case then, we might expect to have to offer a risk premium to entice them to invest for two years and therefore:

$$(1 + r_1) < \frac{(1 + y_2)^2}{1 + E(r_2)}$$

Rearranging terms and simplifying we can see that:

$$E(r_2) < f_2$$

Now notice that in the case where investors wanted to invest long term $E(r_2) > f_2$ but in the case where investors wanted to invest short term $E(r_2) < f_2$. What this shows you is that the relationship between the forward rate and the expected future short rate depends on the relative preferences for long term and short term bonds and investors' willingness to assume interest rate risk.

Theories of the Term Structure

The discussion so far suggested that spot rates for different time periods might differ from each other. This relationship between the various spot rates is known as the *term structure.* If spot rates are higher for longer time periods, then we say the term structure slopes upward, if the rates are all the same we say it is flat, and if the rates are higher for earlier time periods we say it is downward sloping. Now we need to address why the term structure may have these various shapes.

Expectations Hypothesis

This argues that what investors care about is their expected yield from various investments. Using the logic we went through before with respect to forward rates and expected future short rates, this theory simply says that:

$$E(r_2) = f_2$$

This means that we can express the relationship between the spot rates as:

$$(1 + r_1)[1 + E(r_2)] = (1 + y_2)^2$$

Therefore, if the two year rate (y_2) is higher than the one year rate ($r_1 = y_1$), meaning if the term structure is upward sloping, then it must be because we expect the short rate to rise. Similarly, if the term structure is downward sloping, it must be because we expect the short rate to fall.

Liquidity Preference Theory

The discussion above assumed that investors were concerned with the expected outcome only. We know though that people also care about risk and that they require risk premiums to assume risk.

As we showed earlier, for someone who wants to invest for two years, investing at the two year spot rate has no risk but investing for one year and reinvesting for another year does have risk because we cannot be certain that the spot rate one year from now will in fact equal its expected value. Therefore, they will want the rollover strategy to have a higher expected outcome to compensate for this risk and therefore they will demand a risk premium to invest short term. Similarly, short term investors will have to be enticed to invest long term and they will demand a risk premium to invest long term.

So if this theory is correct, the term structure does not reflect expected future spot rates but instead merely reflects whether there exists more long term or short term investors in the market. If there are more short term investors, then the term structure will slope upwards reflecting a risk premium for investing long term and if there are more long term investors then the term structure will slope downwards reflecting a risk premium for investing short term.

The connection between the two alternate approaches presented is that the expectations hypothesis assumes that the forward rate is equal to the expected spot rate whereas the liquidity preference theory implies that the forward rate is equal to the expected spot rate plus a liquidity premium, which could be either positive or negative.

Segmentation Theory

In Hull, there is also a discussion of the Segmentation Theory. This theory of the term structure argues that different borrowers and lenders have different preferences for short-, medium- or long-term investments and they do not readily switch from one maturity range to another. The yields in these different markets are determined by supply and demand in each of these markets.

Interpreting the Term Structure

As we discussed earlier with respect to the Expectations Hypothesis, if investors believe that future spot rates will be higher than the current spot rates, the yield curve will be upward sloping. But we also saw that according to the Liquidity Premium Theory that investors likely place a liquidity premium (could be either positive or negative) on long term rates and thus if the liquidity premium is positive then the yield curve will be upward sloping even if expected spot rates are constant. Therefore, it is impossible to accurately back into the market's expected future short rates from the shape of the yield curve alone. This problem is made worse by the fact that liquidity premiums are unlikely to be constant.

One exception to the above statement is that arguably liquidity premiums tend to be positive and therefore a downward sloping yield curve might be an indication of expected declines in interest rates, either because of expected declines in the real rate or in the inflation rate.

Forward Rates as Forward Contracts

The forward rate discussed earlier was presented as a notional amount that we can calculate and refer to but it was not necessarily a rate that one could lock in. You could actually do this with a forward contract (discussed in Hull), but you can also do this synthetically using zero coupon bonds.

Suppose that the one-year annual yield was 8% and the two year annual yield was 9%. The one year forward rate can be found to be:

$$f_2 = \frac{1.09^2}{1.08} - 1 = 10.01\%$$

Now suppose you wanted to actually lock in the 10% rate as the interest you will pay on a one-year loan in one year. You want to receive \$1,000 in one year and pay back \$1,000(1.1001) = \$1,100.1 in two years.

You can do this with zero coupon bonds by simply buying (investing in) a one-year bond and selling (borrowing) a two-year bond in amounts such that the immediate cash flow is zero. Since the one year zero coupon bond costs \$1,000/1.08 = \$925.93 and the two year zero coupon bond costs \$1,000/1.09^2 = \$841.68, you could simply sell 1.1001 two year bonds, for total proceeds of \$925.93 and at the same time buy a one-year bond. The net up-front payment is zero. In one year, you will get \$1,000 back for your one year bond and then in two years you will repay \$1,000 for each 2-year bond you sold, or 1.1001(\$1,000) = \$1,100.1 in total.

This exactly matches the cash flows you were looking for and is essentially the same as borrowing at the forward rate.

Practice Questions

Question 1. Assume the one-year spot rate is 9% and the two-year spot rate is 10%. What is the forward rate from year one to year 2?

Solution.

$$(1.09)(1 + f_2) = 1.10^2 \Rightarrow f_2 = 11.01\%$$

Question 2. Assume that the expectations hypothesis of the term structure is true and that the one-year spot rate is 10%. If the two-year spot rate is 9%, what is the expected one-year rate one year from now? What is the forward rate?

Solution. If the expectations hypothesis is true, then the expected value of investing for one year and then reinvesting for another year must equal the value of investing for two years at the two-year spot rate. So,

$$(1.10)(1 + E(r_2)) = 1.09^2 \Rightarrow E(r_2) = 8.01\%$$

Note that this also equals the forward rate.

Question 3. Assume that the term structure is upward sloping. Which has a higher yield, a zero coupon bond or a bond which pays a coupon and has the same maturity date?

Solution. The yield on a bond represents an average of the various spot rates, so a coupon bond will have a yield that reflects at least in part the lower spot rates in the early periods and therefore the yield will be lower than for a zero coupon bond.

Question 4. Assume the term structure is downward sloping and that the one-year spot rate is 10% and the two-year spot rate is 5%. Which has a higher yield, a 4% coupon bond or a 12% coupon bond? Assume both bonds pay coupons annually and mature in 2 years.

Solution. You could do the math, but try it with logic. Remember, yields are an average of the spot rates for the various cash flows. Now look at the cash flows for the two bonds: 4 and 104 for the 4% bond and 12 and 112 for the 12% bond.

The bond with the higher coupon has comparatively more of its flows coming earlier, so its yield will be closer to the one-year spot rate relative to the bond with lower coupons, which have relatively more of their cash coming at maturity. Since the term structure slopes downwards, the yield for the 12% bond will be higher.

Question 5. The text discusses a bootstrapping method to calculate spot rates from U.S. Government coupon bonds. Use this method to determine the 6-month, 1-year and 1.5-year spot rates on a bond equivalent basis. Recall that the bond-equivalent basis just means that coupons are paid twice per year and the yield is quoted by multiplying the half-year yield by two.

Assume the following information depicts the annual bond equivalent yields on U.S. Treasury bonds, each with a face value of $1,000. The bond with .5 years to maturity is a zero coupon bond and the others are coupon bonds that trade at par (i.e. price equals face value) with semi-annual coupons.

TABLE 1. Bond Equivalent U.S. Treasury Yields

Maturity (years)	Yield
0.5	2.16%
1.0	2.76%
1.5	3.25%

Rounding could cause a problem for you here, so be sure to carry all calculations to 5 decimal points.

Solution. As stated in the question, the 1 year bond will cost \$1,000 (par). What are its cash flows? The question stated the yield but did not specify what the coupon rate was. But remember that bonds trade at par when the coupons and the yields are equal. Which means, that the coupon rate for the 1-year bond must be 2.76%, its yield, and its cash flows must be $(2.76\%/2)(\$1{,}000) = \$13.80$ in six months and \$1,013.8 in 1 year. Similarly, the 1.5 year bond must have coupons of 3.25% and cash flows of \$16.25 in 6 months, \$16.25 in 1 year and \$1,016.25 in 1.5 years.

Now, to use the bootstrapping technique, we start with the 6 month bond and note that it only has a single cash flow and therefore its yield is in fact a spot rate for $t = .5$. Denote that rate as $y_{.5} = 2.16\%$ on a bond equivalent basis.

Next, from the information we have for the 1 year bond, we know that its price is \$1,000 and its cash flows are \$13.80 and \$1,013.80. Since we know $y_{.5} = 2.16\%$, we simply need to solve the following equation for the unknown quantity y_1:

$$\$1{,}000 = \frac{13.80}{(1 + y_{.5}/2)^1} + \frac{1{,}013.80}{(1 + y_1/2)^2}$$

Solving for y_1 we get $y_1 = 2.7642\%$. Notice that this is .42 basis points higher than the yield given for that maturity on a coupon bond basis.

Continuing on again, we now can use the 6 month and 1 year spot rates to solve for the 1.5 year spot rate. Here, using the price and cash flows for the 1.5 year bond:

$$\$1{,}000 = \frac{16.25}{(1 + y_{.5}/2)^1} + \frac{16.25}{(1 + y_1/2)^2} + \frac{1016.25}{(1 + y_{1.5}/2)^3}$$

Then plug in the values we found for $y_{.5}$ and y_1 to get $y_{1.5} = 3.2613\%$. Notice that this is 1.13 basis points higher than the yield given for that maturity on a coupon bond basis.

Note — The par coupon rates used in this question are published by the U.S. Treasury daily and are referred to as Constant Maturity Treasury (CMT) yields, though only selected maturity dates are shown. The 1.5 year yield in the question was linearly interpolated from the 1 and 2 year CMT yields.

BKM Sections 16.1 and 16.2: Duration and Convexity

Note that this section of the notes covers only Sections 16.1 and 16.2 of the BKM text. It is separated from the rest of Chapter 16 simply so that the similar material in Hull Chapter 4 can be covered before getting into the Asset-Liability Management applications covered in the rest of BKM Chapter 16.

Interest Rate Sensitivity

The first part of this chapter focuses on the sensitivity of bond prices to changes in yields. The text provides an example of four different bonds with different maturities, coupons and initial yields and then examines the changes in price as the yields change.

The four bonds that are used have the following terms:

TABLE 1. Bond Assumptions

Bond	Coupon	Maturity (Years)	Initial YTM
A	12%	5	10%
B	12%	30	10%
C	3%	30	10%
D	3%	30	6%

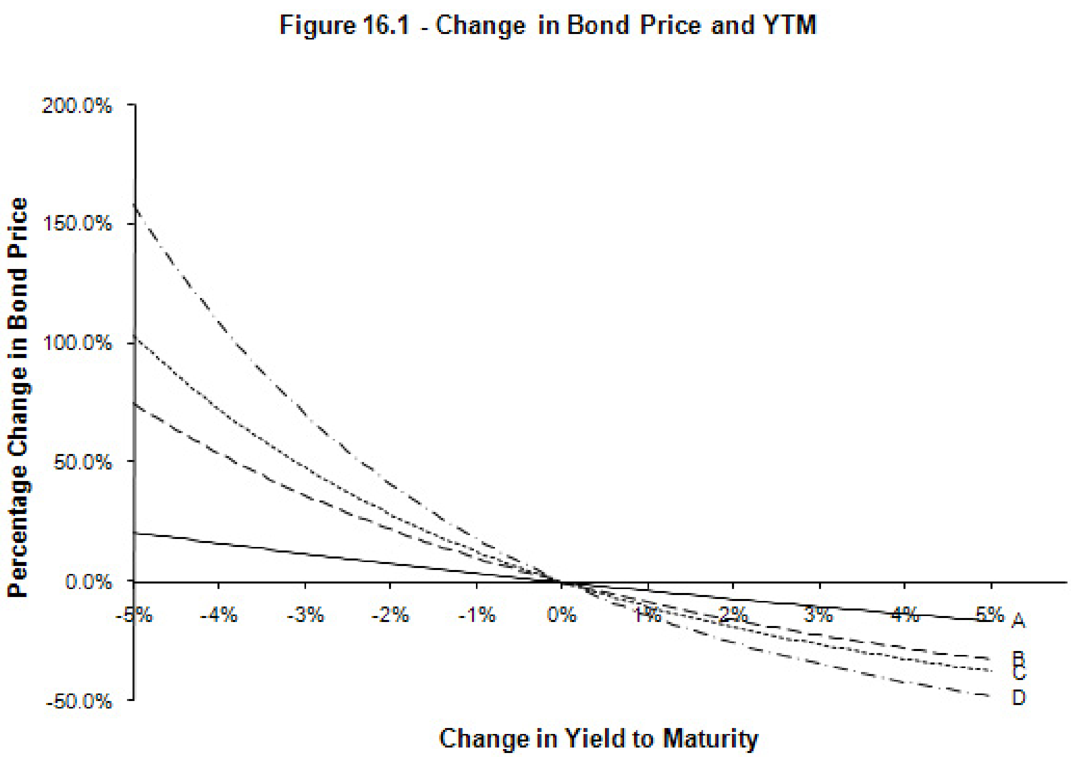
The graph they show as Figure 16.1 is reproduced as Figure 1 on the following page. Using this graph, they specify six general properties of bond prices:

1. Bond prices and yields are inversely related.
2. Increases in yield to maturity has a smaller impact than a comparable decrease in yield.
3. Long term bonds are more sensitive to changes in yields than short term bonds.
4. Sensitivity to changes in yields increases at a decreasing rate as maturity increases.
5. High coupon bonds are less sensitive to changes in yields than low coupon bonds.
6. The sensitivity of a bond's price to changes in yields is inversely related to the yield at which it is currently selling.

Macaulay Duration

A bond's average maturity is the weighted average maturity of each of its cash flows, with the weight given to each cash flow represented by the present value of that cash flow relative to the total bond price. We call this average maturity the *Macaulay Duration, D*.

FIGURE 1. Change in Bond Price



Mathematically, this is simply:

$$D = \frac{(1)CF_1(1+y)^{-1} + (2)CF_2(1+y)^{-2} + \cdots + (n)CF_n(1+y)^{-n}}{\text{Bond Price}}$$

Macaulay Duration represents the weighted average of the time until receipt of each payment. The time to payment is given by the $1, 2, \ldots, n$ terms and the weights are given by:

$$W_t = \frac{CF_t(1+y)^{-t}}{\text{Bond Price}}$$

which represent the proportion of the bond value that each cash flow represents.

Percentage Change in Price

One reason why Macaulay Duration, D, is so important is that the percentage change in price for a bond for a given proportional change in the bond's yield is directly related to its duration. Mathematically, this can be written as:

$$\frac{\Delta P}{P} = -D\left[\frac{\Delta(1+y)}{1+y}\right]$$

In practice, it is common to define *Modified Duration* as the Macaulay Duration divided by 1 plus the yield, or:

$$D^* = \frac{D}{1+y}$$

Then noting that $\Delta(1+y) = \Delta y$, we can rewrite the percentage change in price as:

$$\frac{\Delta P}{P} = -D^* \, \Delta y$$

Modified Duration

To better understand the relationship between modified duration and the percentage change in price, we need to understand why modified duration measures price sensitivity. The modified duration is related to the first derivative of the bond price function with respect to the interest rate.

Modified Duration with Annually Compounded Yields

Consider a simple case of a bond that pays annual cash flows of CF_i for n periods. The price of the bond is determined by discounting the cash flows at the annually compounded yield, y:

$$P = CF_1(1+y)^{-1} + CF_2(1+y)^{-2} + \cdots + CF_n(1+y)^{-n}$$

Suppose the yield changes by a small amount. The derivative of the price function with respect to the yield tells us the change in the price of the bond per unit change in the yield. This derivative is easily found as:

$$\begin{aligned} \frac{dP}{dy} &= (-1)CF_1(1+y)^{-2} + (-2)CF_2(1+y)^{-3} + \cdots + (-n)CF_n(1+y)^{-n-1} \\ &= \frac{-1}{1+y}\left[(1)CF_1(1+y)^{-1} + (2)CF_2(1+y)^{-2} + \cdots + (n)CF_n(1+y)^{-n}\right] \end{aligned}$$

We can divide the whole equation above by the current price of the bond to present this in terms of a percentage change in price.

$$\%\ \text{Price Change} = \frac{-1}{1+y}\left[\frac{(1)CF_1(1+y)^{-1} + \cdots + (n)CF_n(1+y)^{-n}}{\text{Bond Price}}\right]$$

Simplifying, we can define the term in the brackets as the Macaulay Duration, D, and then the change in price as a percent of initial price is:

$$\frac{dP}{P} = \frac{-1}{1+y} D \, dy$$

Note that the terminology can be confusing since we just defined two types of duration.

- We used the term *Macaulay Duration* to represent the quantity:

$$D = \frac{(1)CF_1(1+y)^{-1} + \cdots (n)CF_n(1+y)^{-n}}{\text{Bond Price}}$$

  This is nothing more than the weighted average time to payment.

- *Modified Duration* is similar to this quantity but also includes the $1/(1+y)$ term. The relationship between the two is simply:

$$\text{Modified Duration} = \frac{\text{Macaulay Duration}}{1+y}$$

$$D^* = \frac{D}{1+y}$$

Shortcut Formula for Modified Duration

Note that it can be time consuming to compute modified duration by hand for a long maturity bond. However, it can be approximated very easily with your financial calculators. Simply define P as the current price of the bond, then calculate the price if the yield goes up by a small amount $\Delta y = .1\%$ or down by a small amount $\Delta y = -.1\%$ and define P_- as the price when yields go down and P_+ as the price when yields go up.

Then an approximation for Modified Duration is given as:

$$\text{Modified Duration} \approx \frac{P_- - P_+}{2P\Delta y}$$

To see the approximation more clearly, note in Figure 2 on the next page that we want to measure the slope of the tangent line at the current yield, which is 7% in this case. If we use the points P_- and P_+ as *approximations* of points on the tangent line, then we can approximate the slope of the tangent line as the rise over the run. The rise is the difference between P_- and P_+ and the run is twice the length of Δy. The resulting ratio approximates the derivative of the price function:

$$\frac{dP}{dy} \approx \frac{P_- - P_+}{2\Delta y}$$

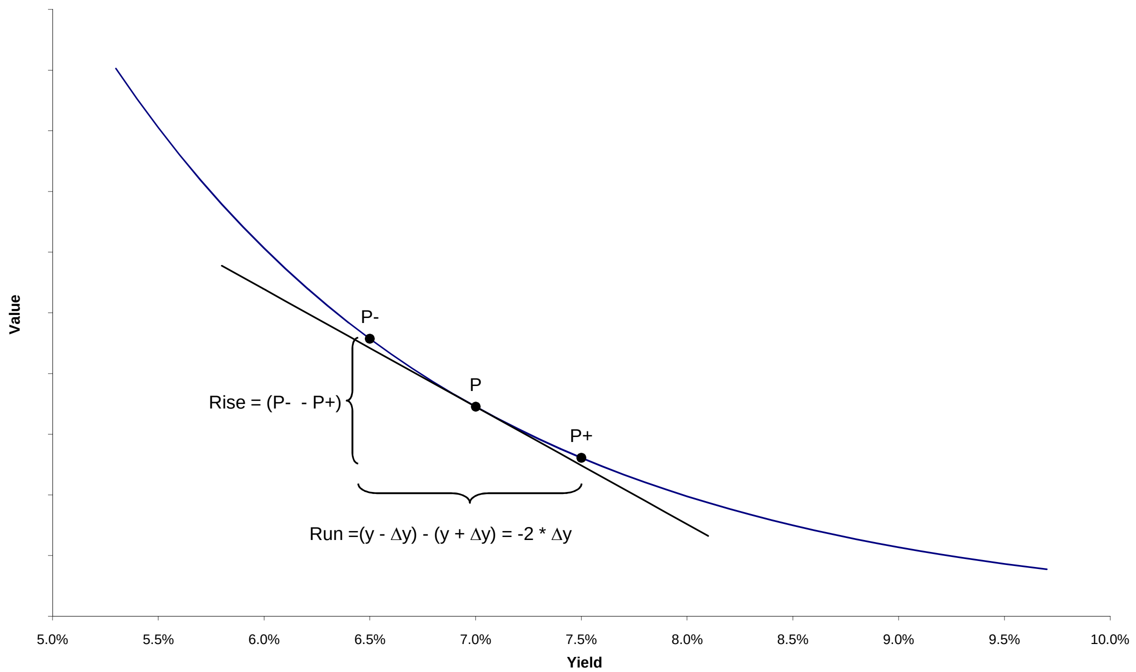
If we further express the derivative as a percentage of the current price, we can divide by P and the formula for the Modified Duration results:

$$\text{Modified Duration} \approx \frac{P_- - P_+}{2P\Delta y}$$

Modified Duration with Continuous Compounding

The modified duration formula with continuous compounding is defined in exactly the same way as before, as the first derivative of the value of the bond for a small change in the yield, expressed relative to the initial bond price and ignoring the negative sign. With continuous

FIGURE 2. Approximating Duration



compounding, the bond value is given as:

$$P = \sum c_i e^{-yt_i}$$

and then the derivative is simply:

$$\frac{\partial P}{\partial y} = -\sum t_i c_i e^{-yt_i}$$

dividing by the initial bond value, P, and ignoring the negative sign, duration is then defined as:

$$D = \frac{\sum t_i c_i e^{-yt_i}}{P}$$

Notice that with continuous compounding, the Macaulay and Modified Durations are equal and both can be expressed as D.

Modified Duration with Semi-Annual Compounding

Notice that aside from the mechanical differences in calculating the present values, the only important difference between the modified duration formulas with annual compounding and with continuous compounding is the term $1/(1+y)$ that is included in the case of annual compounding. That is, the Macaulay duration is calculated in the same manner as the weighted average time to payment and only the modified duration formula differs.

Similarly, when the bond values are calculated using semi-annually compounded rates, this term is adjusted to be $1/(1 + y/2)$, as shown below:

$$\text{Modified Duration} = \frac{\text{Macaulay Duration}}{1 + y/2}$$

More generally, if the compounding frequency is k times per year, the denominator would simply be $(1 + y/k)$, which explains why when compounding is continuous such that $k = \infty$, the denominator approaches 1.0 and modified duration is simply equal to Macaulay duration.

Explanation for the Relationship Between the Price Change and the Modified Duration

Having defined the modified duration in terms of the derivative of the price with respect to the change in the yield, it is helpful to demonstrate more fully the relationship between the price change and the yield change in terms of a Taylor expansion.

Taylor Expansion

If you have a function $f(x)$ and want to know the value of this function at a point $x + h$, then a Taylor Expansion suggests that:

$$f(x + h) = f(x) + f'(x)\Delta x + \frac{1}{2}f''(x)(\Delta x)^2 + \boldsymbol{O}$$

where $\boldsymbol{O}$ represents all of the higher order terms with $(\Delta x)^3$, $(\Delta x)^4$, etc.

When Δx is small we can usually ignore all of the higher order terms since $(\Delta x)^2$ will be *very* small, $(\Delta x)^3$ will be even smaller, etc.

So for small values of Δx, we can ignore all of the terms of order $(\Delta x)^2$ or higher and rewrite the change in the value of the function as:

$$\Delta f = f(x + h) - f(x) \approx f'(x)\Delta x$$

Application to Bond Price Changes

Now apply this logic to the bond price function and changes in the yield. Here we could write the change in the value of the bond as:

$$\Delta P = P'\Delta y$$

Now, if we simply divide both sides by the current bond price, P, and note that the first derivative divided by the price was close to how we defined duration, except for the fact that we multiplied by negative one, then we can rewrite this as:

$$\frac{\Delta P}{P} = -\text{Modified Duration}\,\Delta y$$

Convexity Adjustment

Duration is only good as an approximation of the price sensitivity for small changes in rates. This is because the curvature of the price function makes the estimate based solely on the first derivative (the tangent line) a bad estimate for large changes. To put it in terms of the Taylor expansion, for larger changes in the interest rates it is not appropriate to ignore the second term of the Taylor expansion.

In this case, the change in the bond price has to be written as:

$$\Delta P = P' \Delta y + \frac{1}{2} P'' (\Delta y)^2$$

To reflect the second derivative, or the curvature, we define Convexity as the second derivative of the bond price function, P'' in the equation above, divided by the price.

Then this can be written in terms of both duration and convexity as:

$$\frac{\Delta P}{P} = -\text{Modified Duration}\,(\Delta y) + \frac{1}{2}\text{Convexity}\,(\Delta y)^2$$

Formulas for Convexity

The formulas for convexity are similar to the formulas for duration.

Convexity for Bond with Annual Compounding

For a bond with annual cash flows, convexity is calculated as follows:

$$\text{Convextity} = \frac{1}{P(1+y)^2} \sum \left[\frac{CF_t}{(1+y)^t} t(t+1) \right]$$

Note that this formula looks like a mess to remember, but it really is nothing more than the second derivative of the price function with respect to the yield, so it is really easy to derive if you need to.

Convexity with Continuous Compounding

Recall that we defined convexity as the second derivative of the value of the bond. With continuous compounding, and when expressed as a percentage of the bond value, this is simply:

$$C = \frac{1}{P} \frac{\partial^2 P}{\partial y^2} = \frac{\sum t_i^2 c_i e^{-y t_i}}{P}$$

Convexity with Semi-Annual Compounding

As in the case with duration, the convexity formula is just a bit messier if the compounding frequency is k times per year. This isn't covered in the syllabus readings, so rather than worry about the details just note that you can always use the approximation formula below if you need to calculate the convexity for bonds with compounding frequencies other than annual or continuous.

Shortcut Formula for Convexity (Approximation)

It is quite easy to approximate Convexity using your calculators and avoid the time consuming calculations shown above. Similar to the shortcut formula I gave you earlier for duration, Convexity can be approximated as:

$$\text{Convexity} = \frac{P_- + P_+ - 2P}{P(\Delta y)^2}$$

Note that to derive this approximation for convexity, I am simply using the duration approximation discussed earlier at two different points (but note carefully that the two points I am going to use here are the prices at $y + .5\Delta y$ and at $y - .5\Delta y$ and I am using yield changes equal to $.5\Delta y$).

This would give us two estimates of duration:

$$D_{y-.5\Delta y} = \frac{P_{y-.5\Delta y} - P}{P\Delta y} \qquad D_{y+.5\Delta y} = \frac{P - P_{y+.5\Delta y}}{P\Delta y}$$

Then, convexity is just the change in the durations divided by the change in the yields at which the durations were measured. Taking the differences in these two formulas and dividing by the change in the yield, the convexity approximation becomes:

$$\text{Convexity} = \frac{P_- + P_+ - 2P}{P(\Delta y)^2}$$

This can be seen graphically in Figure 3 on the facing page.

Why Do Investors Like Convexity?

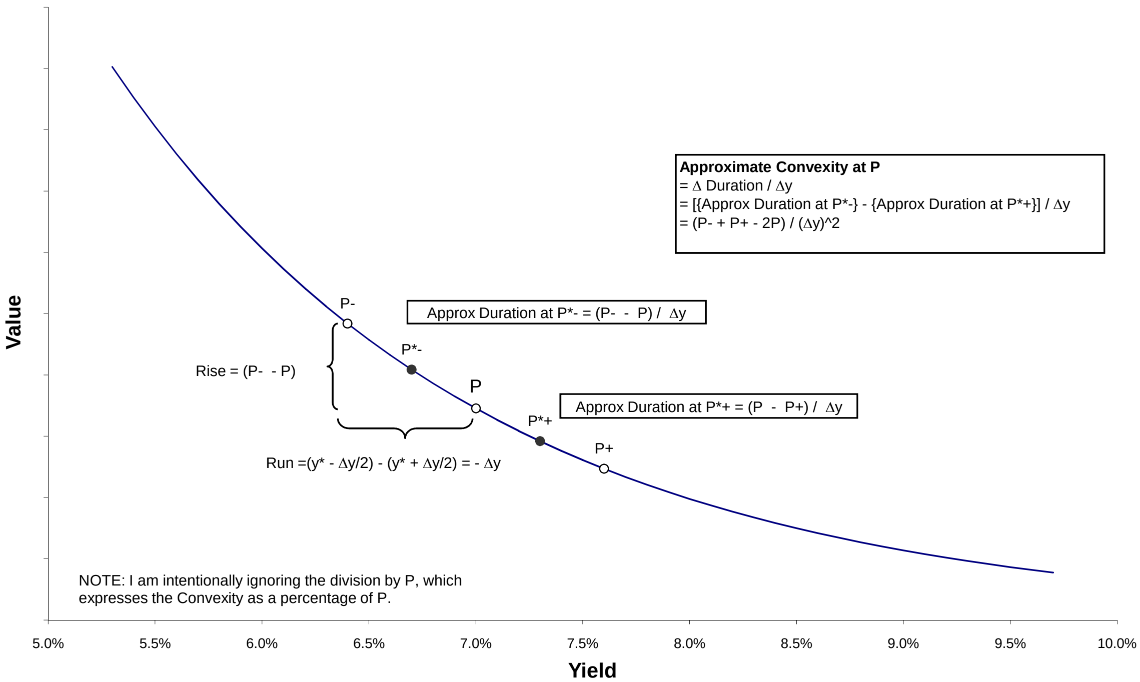
Notice that convexity causes the price of bonds to increase more when yields fall than they decrease when yields rise. This asymmetry is desirable and causes an increase in the expected return for bonds as the yield volatility increases.

Effective Duration and Convexity

One thing you may have noticed is that both the duration and convexity formulas we've discussed assumed that the cash flows were fixed and were not affected by the changes in the yields. In reality, many bonds have cash flows that are a function of the yields and therefore this should be reflected in any duration or convexity calculation. When we reflect the possible changes in the cash flows in our analysis, we refer to the resulting durations and convexity values as the effective duration and effective convexity.

One example of when this would be important is when the bond has a call feature. When yields fall, normally bond values rise. If there is a call feature, when rates fall the issuer is likely to call the bond and any potential price appreciation will be lost.

FIGURE 3. Approximating Convexity



Writing a specific formula for effective duration is a lot more complicated than for Macaulay or Modified duration. Instead, we make use of the *approximations* provided earlier and calculate the price of the bond if interest rates rise (P_+) and the price if rates fall (P_-). In both cases we take into account not just the change in rates but also the change, if any, in the cash flows. It is this explicit consideration for the changes in the cash flows themselves that makes the resulting duration and convexity estimates *effective* duration and *effective* convexity.

Then we measure duration and convexity using the formulas:

$$\text{Effective Duration} = \frac{P_- - P_+}{2P\Delta y}$$

$$\text{Effective Convexity} = \frac{P_- + P_+ - 2P}{P(\Delta y)^2}$$

Notice that the effective duration formula above is not exactly what is shown in the BKM textbook, but that is simply because in the book they use $\Delta P = P_+ - P_-$ and $\Delta r = 2\Delta y$ and then write it as:

$$\text{Effective Duration} = -\frac{\Delta P}{P\Delta r}$$

To a large extent, this is just a minor change in notation, but there are two subtle points to keep in mind. First, the formulas for P+ and P- explicitly take into account all of the factors that would affect the price, including changes to the cash flows, changes in volatility, etc. Second, the price change is measured relative to a shift in the term structure rather than the yield on the bond, which explains why they use the Δr notation.

Mortgage Backed Securities

Mortgage backed securities have interesting durations and convexities that are important to understand.

Because prepayments on mortgages are linked to changes in interest rates, the interest rate sensitivity (duration and convexity) of mortgage backed securities can be quite different from that of a typical bond. They typically exhibit negative convexity, similar to callable bonds, because of the embedded prepayment option. As rates fall, the value of a MBS will not rise as much as in the case of a conventional bond because of the likelihood that the mortgage will be repaid early.

Practice Questions

Question 1. Assume you have a 2-year, 5% coupon bond with annual coupons and a \$1,000 face value. Determine the duration of the bond assuming the current yield is 5% on an annually compounded basis.

Solution. The Macaulay duration is just the weighted average time to each payment, with the weights equal to the PV of each cash flow as a percent of the total price. As shown below, the bond price is \$1,000 and the Macaulay duration is 1.952. Modified duration is then 1.952/1.05 = 1.859.

Period	T	Cash Flow	PV(CF)	T*PV(CF)
1	1	50	47.62	47.62
2	2	1,050	952.38	1,904.76
			1,000.00	1,952.38
			Macaulay Duration	**1.952**
			Modified Duration	**1.859**

Question 2. Approximate the duration of the bond in the previous question by calculating the price when the yields are 5%, 4.9%, 5.1% and using the duration approximation:

$$\text{Modified Duration} = \frac{P_{-} - P_{+}}{2P\Delta y}$$

Solution. Using your financial calculators, it is easy to show that the bond prices at 5%, 4.9% and 5.1% are given as:

$$P = 1{,}000.00$$
$$P_- = 1{,}001.86$$
$$P_+ = 998.14$$

From this, we can find the duration to be:

$$\begin{aligned}\text{Modified Duration} &= \frac{P_- - P_+}{2P\Delta y}\\ &= \frac{1{,}001.86 - 998.14}{2(1{,}000)(.001)}\\ &= 1.859\end{aligned}$$

Question 3. Redo the previous questions assuming that this is a 5-year bond instead of a 2 year bond. How does the duration compare to the duration for the 2-year bond with the same coupons?

Solution. For the 5 year bond, we have the following calculations:

Period	T	Cash Flow	PV(CF)	T*PV(CF)
1	1	50	47.62	47.62
2	2	50	45.35	90.70
3	3	50	43.19	129.58
4	4	50	41.14	164.54
5	5	1,050	822.70	4,113.51
			1,000.00	4,545.95
			Macaulay Duration	**4.546**
			Modified Duration	**4.329**

Notice that this duration is much higher, meaning that the longer maturity bond is much more sensitive to small changes in the yield.

Question 4. Redo the previous question and this time assume a 5-year bond but increase the coupon to 10%. How does this duration compare to the duration of the 5-year 5% bond?

Solution. With the higher coupons, this calculation becomes:

Period	T	Cash Flow	PV(CF)	T*PV(CF)
1	1	100	95.24	95.24
2	2	100	90.70	181.41
3	3	100	86.38	259.15
4	4	100	82.27	329.08
5	5	1,100	861.88	4,309.39
			1216.47	5,174.27
			Macaulay Duration	**4.253**
			Modified Duration	**4.051**

Notice that this duration is lower because with the higher coupons proportionately more of the cash flows come earlier and we already showed previously that shorter maturity cash flows are less sensitive to interest rate changes.

Question 5. Assume that you had a portfolio consisting of each of the three bonds in the previous questions (the 2-year 5% bond, the 5-year 5% bond and the 5-year 10% bond). If the current interest rates are 5%, determine the duration of the portfolio.

Solution. Since we already know the individual durations and the bond values, we simply need to calculate the weighted average of the individual durations. Using the market values of each bond and their durations from the previous questions, we have:

$$\text{Portfolio Duration} = \frac{1.859(1{,}000) + 4.329(1{,}000) + 4.051(1{,}216.47)}{3{,}216.47} = 3.45$$

Question 6. What is the duration of a bond with a face value of \$1,000, 5% coupons, paid annually, and 4 years to maturity if the yield is 4% on an annually compounded basis? Do this using both the full duration formulas and the approximation method presented in the text.

Solution. Using the same formulas we've been using,

Period	T	Cash Flow	PV(CF)	T*PV(CF)
1	1	50	48.08	48.08
2	2	50	46.23	92.46
3	3	50	44.45	133.35
4	4	1,050	897.54	3,590.18
			1,036.30	3,864.06
			Macaulay Duration	**3.729**
			Modified Duration	**3.585**

To use the shortcut formula, let's assume a 20 basis point change in the yield. Using your financial calculators, it is easy to show that the bond prices at yields of 4%, 3.8% and 4.2% are given as:

$$\begin{aligned} P &= 1{,}036.30 \\ P_- &= 1{,}043.76 \\ P_+ &= 1{,}028.90 \end{aligned}$$

From this, we can find the duration to be:

$$\begin{aligned} \text{Modified Duration} &= \frac{P_- - P_+}{2P\Delta y} \\ &= \frac{1{,}043.76 - 1{,}028.90}{2(1{,}036.30)(.002)} \\ &= 3.585 \end{aligned}$$

Question 7. Assume the yield of the bond in the previous question decreases by .1%, to 3.9%. What is the percentage change in the price of the bond? Compare the actual price change to what could be estimated based on its duration.

Solution. The price at the revised yield is found by just changing the yield and discounting the same cash flows to get $P = 1{,}040.02$, which is \$3.72 higher than the original price. This is an increase of $3.72/1{,}036.30 = .359$ percent.

We could get approximately the same answer by noting that the percentage price change is equal to:

$$\text{Approx. \% Change in Price} = -D^* \, \Delta y = (-3.585)(-.001) = .3585$$

Question 8. What is the convexity of the bond in Question 6?

Solution. Using the convexity formula shown in the text,

$$\text{Convextity} = \frac{1}{P(1+y)^2} \sum \left[\frac{CF_t}{(1+y)^t} t(t+1) \right]$$

And the specific calculations are as follows:

Period	T	Cash Flow	PV(CF)	(T+1)*T*PV(CF)
1	1	50	48.08	96.15
2	2	50	46.23	277.37
3	3	50	44.45	533.40
4	4	1,050	897.54	17,950.89
			1,036.30	18,857.81
			Convexity	**16.824**

We could also use the approximation formula similar to the one we used for the duration calculation. Recall that the bond prices at yields of 4%, 3.8% and 4.2% were:

$$P = 1{,}036.30$$

$$P_{-} = 1{,}043.76$$

$$P_{+} = 1{,}028.90$$

Being very careful to carry all of the decimal places, we can approximate convexity as:

$$\begin{aligned}\text{Convexity} &= \frac{P_{-} + P_{+} - 2P}{P(\Delta y)^2} \\ &= \frac{1{,}028.90281 + 1{,}043.76484 - 2(1{,}036.2989)}{1{,}036.2989(.002)^2} \\ &= 16.827\end{aligned}$$

Question 9. Determine the actual percentage price change of the bond in the previous question if the yield increases .5% to 4.5% and compare that to the estimate based upon Duration. Then use the convexity adjustment and the convexity measured previously to improve the estimate of the price change based on the duration (again, be careful not to round).

Solution. The price at 4.5% will be \$1,017.9376, which is a change of −1.7718%. Based on duration, the price change would be estimated to equal:

$$\text{Approximate Price Change} = -3.585(.005) = -1.7926\%$$

If we also include the convexity adjustment, we would have to add in the amount equal to:

$$\begin{aligned}\text{Convexity Adjustment} &= \frac{1}{2}\text{Convexity}\,(\Delta y)^2 \\ &= \frac{1}{2}(16.824)(.005^2) \\ &= .02103\end{aligned}$$

The total price change is then:

$$\text{Price Change} = -1.7926\% + .02103\% = -1.7716\%$$

This is very close to the actual change of -1.7718%. As you can see, for a small change in yield duration alone is pretty close. For larger changes in yield, it is important to include the convexity adjustment.

Hull Chapter 4: Interest Rates

This chapter contains a lot of material that is also covered in the BKM text, specifically determining zero rates and forward rates and calculating duration and convexity. However, the notation in Hull is different because of the use of continuous compounding. The unique elements in this chapter to pay attention to are the duration and convexity formulas with continuously compounded interest rates and the discussion of Forward Rate Agreements.

Interest Rates

Treasury rates are the interest rates paid on U.S. government debt instruments including bills, notes and bonds. They are generally considered to be free of default risk and are commonly used to represent the risk-free rate of interest.

However, while many investors can invest at these risk-free rates by buying U.S. government debt, this isn't a rate at which entities other than the U.S. government can borrow money. Most financial instruments we care about for the present purposes (swaps, derivatives, etc.) involve transactions among non-government entities. So what we really care about is the rate at which institutions such as large, highly-rated (but not literally risk-free) banks can borrow or lend.

Prior to the 2007–2009 credit crisis, the most common rate to use for this purpose was the London Interbank Offered Rate, or LIBOR. It is a reference rate, published daily, designed to reflect the rate at which large banks are willing to lend to other banks. Since the credit crisis, it is now more common to use the Overnight Indexed Swap rate (OIS) as the "risk-free rate".

Zero Rates

BKM Chapter 15 discussed how to determine the zero coupon yield curve using the prices of coupon bonds. This is the same material discussed in this reading, but presented using continuous compounding. The numerical problems will highlight the practical differences, but the approach is identical to what was already presented.

Bond Pricing

This chapter contains background information on Bond Pricing, which I covered at the start of this section of the study manual.

For the exam, and to follow the remaining readings, you should be able to use continuously compounded interest rates to value a bond and to calculate its yield.

$$P = \sum CF_t e^{-yt}$$

You will have ample opportunities to review these calculations in the discussion of interest rate swaps in Hull Chapter 7.

Nonetheless, it is worth pointing out here the important definition of the *par yield*, which is also often referred to as the *swap rate*. This refers to the coupon rate on a semi-annual bond whose price is exactly equal to its par value.

Determining Zero Rates

The text presents a bootstrapping method to sequentially determine the zero coupon rates from coupon bond prices. This is the same as what was done in BKM, but in a somewhat simplified manner. In addition, because Hull uses continuous compounding, the formulas are a bit different but the concept is the same.

Start by assuming you know the price P of a six-month zero coupon bond with a principal of \$100. The 6 month zero coupon rate on a continuous basis is simply the rate R_1 which solves the equation:

$$P = 100e^{-R_1(.5)}$$

Note that R_1 is a rate for a 6-month cash flow but it is stated on an annual basis.

Now, suppose we didn't have a 1-year zero coupon bond, but we did have a 1-year coupon bond with a coupon of \$5 in 6 months and principal and interest of \$105 in 12 months. If the price of this is P, then we get the 1-year zero coupon rate by solving the equation:

$$P = 5e^{-R_1(.5)} + 105e^{-R_2(1)}$$

Forward Rates

Once we determined the zero rates in BKM, we found the forward rates by noting that:

$$(1 + y_2)^2 = (1 + y_1)(1 + f_2)$$

With continuous compounding, and denoting the 1-year, 2-year and forward rates as R_1, R_2 and R_F respectively, this becomes:

$$e^{R_2(2)} = e^{R_1(1)}e^{R_F(1)}$$

Then, we just solve for R_F, which in this case is very easy,

$$R_F = \frac{R_2(2) - R_1(1)}{1}$$

Using T_i to indicate the time periods, this can also be written more generally as:

$$R_F = \frac{R_2(T_2) - R_1(T_1)}{T_2 - T_1}$$

Notice that this fraction can also be written as:

$$R_F = R_2 + (R_2 - R_1)\frac{(T_1)}{T_2 - T_1}$$

Using this form, we can see that if we reduce the time between periods, we can write the formula for the *instantaneous forward rate* as:

$$R_F = R + T\frac{\partial R}{\partial T}$$

Forward Rate Agreement (FRA)

A *Forward Rate Agreement* is simply a contract in which one party agrees that for some time period in the future he will pay an interest rate of R_K (at a specified compounding frequency) on a specified principal amount, L. At time T_1 the person with the long position will pay L and at T_2 will receive L plus interest equal to $LR_K(T_2 - T_1)$.

Note: In the problems associated with FRAs, the interest payment amounts are usually stated in simple interest terms, with a compounding frequency such as annual, semi-annual or quarterly. However, there will also be the need to determine the present values of FRA payments. These will be determined using continuously compounded rates.

To determine the *value* of an FRA at any point in time, note that we could always earn the forward rate R_F during the period T_1 to T_2 with no up-front cost by borrowing the present value of L now until T_1 and investing the same amount until T_2. Therefore, there's no up-front cash outlay, but at time T_1 you have to repay the borrowing with interest to T_1 (i.e. pay L) but then at T_2 you get the invested amount back with interest to T_2. The net effect is that from T_1 to T_2 the rate you earn is the forward rate on an amount L.

Since you can do that now with no cost, then an FRA with R_K equal to R_F would have no value (something that you can do for free can't have any value).

If $R_K > R_F$ then the value of the FRA would simply be the present value of the amount of additional interest earned,

$$\text{Value of FRA} = L(R_K - R_F)(T_2 - T_1)e^{-R_2 T_2}$$

Consistent with the previous note regarding the compounding frequency of the FRA rate, R_K, in the previous equation usually R_K and R_F will be based on either an annual, semi-annual or quarterly compounding but R_2 will be given as a continuously compounded rate for a cash flow at time T_2.

Keep in mind something subtle with respect to the value of an FRA. We are not talking here about the value of the interest and principal cash flows. We are talking about the value of the *agreement*. If the agreement merely allows you to receive the market rate of interest, you could have done that on your own and therefore the agreement itself has no value. If the agreement allows you to receive a better than market rate of interest, then this has a positive value. This is why the formula above reflects the difference between the agreed upon rate, R_K, and the forward rate, R_F. This difference is then multiplied by the principal amount, adjusted to reflect the time periods and then multiplied by a present value factor.

An Alternative Approach

Note that an FRA requires that you pay L at T_1 and then you receive interest and principal back at T_2. If you were to enter into the FRA and at the same time simply plan to borrow L at T_1 and then repay it at whatever interest rate prevails at that time, R, then there will be no up-front cash flows, no net cash flows at T_1 and cash flows at T_2 will consist of the proceeds from the FRA less the repayment of the borrowings:

$$\begin{aligned}\text{Cash Flows at } T_2 &= [LR_K(T_2 - T_1) + L] - [LR(T_2 - T_1) + L] \\ &= L(R_K - R)(T_2 - T_1)\end{aligned}$$

To value this cash flow now, keep in mind that R is an unknown rate at this point. We won't know what that rate is until time T_1. To determine the value of this cash flow, we would need to know its expected value (in order to know what R may turn out to be) and would also need to know the appropriate discount rate to use (again, since R is not known until T_1, the cash flow is risky and may need to be discounted at a risk-adjusted rate).

This suggests that we could value this cash flow using the expected rate, $E(R)$ in place of the unknown quantity R and discounting at a risk-adjusted rate, R^*:

$$\text{Value of FRA} = L[R_K - E(R)](T_2 - T_1)e^{-R^* T_2}$$

However, from our previous result we know that the correct value is:

$$\text{Value of FRA} = L(R_K - R_F)(T_2 - T_1)e^{-R_2 T_2}$$

So we could get the right value by simply assuming that $E(R) = R_F$ and then discounting at the risk free rate, R_2. This will give us the same answer as before. That is, if we pretend the expected rate is the forward rate then we can discount at the risk-free rate and not have to worry what the right risk-adjusted discount rate might be.

This last step highlights a critical concept in the context of pricing swaps, options and other derivative securities known as *Risk Neutral Pricing*. This "trick" involves altering the probability distribution so that the expected rate is equal to the forward rate, using that to estimate the cash flows and then discounting using a risk-free rate. You should have been exposed to this on the preliminary actuarial exams, including Exam 3F. If not, you will need to pay careful attention to this again in Hull Chapter 7 when we use the technique to value interest rate swaps and currency swaps.

Duration and Convexity

This section simply repeats the material from BKM Chapter 16 regarding duration and convexity. The only thing to note is that Hull using continuous rather than annual compounding. As a result, there is no distinction made between the Macaulay and Modified durations, as was done in the BKM reading. When interest rates are continuously compounded, the Macaulay and Modified durations are identical.

The duration formula is defined in exactly the same way as before — it represents the first derivative of the value of the bond for a small change in the yield, expressed relative to the initial bond price and ignoring the negative sign. With continuous compounding, the bond value is given as:

$$B = \sum c_i e^{-yt_i}$$

and then the derivative is simply:

$$\frac{\partial B}{\partial y} = -\sum t_i c_i e^{-yt_i}$$

dividing by the initial bond value, B, and ignoring the negative sign, duration is then defined as:

$$D = \frac{\sum t_i c_i e^{-yt_i}}{B}$$

Notice that with continuous compounding, duration can still defined as the weighted average time to payment (that's how we defined Macaulay Duration) where the weights are the present values of the cash flows and expressed as a percentage of the total bond value. However, when compounding was annual or semi-annual, the derivative (which we defined as Modified Duration), actually had an additional term of $\frac{1}{1+y/m}$.

For a small change in the yield the change in the value of the bond (as a percentage of the original bond price) can be determined using duration according to the same formula used in BKM:

$$\frac{\partial B}{B} = -D\,\partial y$$

This formula applies only for small changes in the yield. For larger changes, the Taylor expansion can be used to more accurately express the approximate change in the value in terms of both the first derivative (Duration) and the second derivative (Convexity). The previous expression can then be modified as follows:

$$\frac{\partial B}{B} = -D\,\partial y + \frac{1}{2}C(\partial y)^2$$

Recall that BKM defined convexity as the second derivative of the value of the bond. Using continuous compounding and expressing this derivative as a percentage of the bond value, the convexity is simply:

$$C = \frac{1}{B}\frac{\partial^2 B}{\partial y^2} = \frac{\sum t_i^2 c_i e^{-yt_i}}{B}$$

Theories of the Term Structure

This is a brief summary of the material presented in BKM Chapter 15 regarding the Expectations Hypothesis and the Liquidity Preference Theory.

In Hull, there is also a discussion of the Market Segmentation Theory, which argues that different borrowers and lenders have different preferences for short-, medium- or long-term investments and they do not readily switch from one maturity range to another. The yields in these different markets are determined by supply and demand in each of these markets.

Hull makes the point that the Liquidity Preference Theory is the most appealing and consistent with empirical evidence. He uses the example of a bank trying to manage its interest rate risk to show why it makes sense for longer term rates to generally exceed the rates implied by expected future short term rates, and hence exhibit a *liquidity premium.*

Practice Questions

Question 1. Assume an interest rate is quoted as being 10% per annum with quarterly compounding. What is the equivalent rate with continuous compounding?

Solution. With continuous compounding, the terminal value of $1 in n years is determined as e^{rn} and with quarterly compounding it is given as $(1+r/4)^{4n}$. Setting these equal and solving for the continuous rate you get $e^r = (1.025)^4$ which can be solved for $r = 9.877\%$.

Question 2. Assume that the price of a 6-month zero coupon bond with a face value of 100 is 98 and the price of a 1-year coupon bond with 8% coupons paid semi-annually and a face value of 100 is 102. Determine the 6-month and 12-month zero coupon rates on a continuously compounded basis.

Solution. The first step is to use the zero coupon 6-month bond to directly solve for the 6-month zero coupon rate. In this case, setting the price of the bond equal to the present value of its cash flow:

$$98 = 100e^{-R(.5)}$$

and then R can be found easily to be:

$$R = -ln(98/100)(2) = 4.04\%$$

Then, to get the one-year rate, we note that the 1-year bond has cash flows of 4 in 6 months (one half of the coupon rate times the face value) and 104 in 12 months. The price of this bond was given as 102 which means that:

$$102 = 4e^{-.0404(.5)} + 104e^{-R(1)}$$

and then we can solve for $R = 5.86\%$.

Question 3. Assume that the one-year zero rate is 9% and the two-year zero rate is 10%, both continuously compounded. What is the forward rate, continuously compounded, between time periods one and two (the one year rate one year from now)?

Solution. Recall that using annual compounding, the relationship between the forward rate (R_F), the one-year rate (R_1) and the two-year rate (R_2) is as follows:

$$(1 + R_2)^2 = (1 + R_1)(1 + R_F)$$

On a continuous basis, the comparable relationship is:

$$e^{2R_2} = e^{R_1}e^{R_F}$$

From this, it is easy to solve for $R_F = 11\%$.

Note that you can think about this in the following way. If you want to lock in a forward rate from time period 1 to time period 2, you can do this in either of two ways. One, you can enter into a forward agreement to invest L at $T = 1$ and get back Le^{R_F} at $T = 2$, where R_F is whatever rate you agree to today to apply to this investment at $T = 1$ (i.e. the forward rate).

Alternatively, you could first borrow the present value of some amount, L, with the plan to repay it in one year. The amount you will borrow will be Le^{-R_1}. At the same time, you can take that money and invest it for two years, for a payoff at year two of $(Le^{-R_1})e^{2R_2}$.

The net effect is that at $T = 0$ you pay nothing, at $T = 1$ you pay L (the repayment of the borrowed money) and then at $T = 2$ you receive:

$$(Le^{-R_1})e^{2R_2} = Le^{2R_2-R_1}$$

Obviously, since both strategies have no risk and both involve paying L at $T = 1$ and getting some money back at $T = 2$, the money at $T = 2$ must be the same. That tells us that:

$$Le^{2R_2-R_1} = Le^{R_F}$$

dividing by L,

$$e^{2R_2-R_1} = e^{R_F}$$

$$e^{2R_2} = e^{R_1}e^{R_F}$$

Question 4. Using the same facts as in the previous question, suppose that six months ago we had entered into an FRA where we had agreed to receive 12% with annual compounding on a principal amount of \$10 million between time period $T_1 = 1.5$ years and $T_2 = 2.5$ years. What is the value of this FRA as of today (six months after we did the deal)? Assume that the one-year zero rate today is 9% and the two-year zero rate today is 10%, both continuously compounded.

Solution. First, because the FRA is based on annual compounding, we need to determine the forward rate that we could get today on the same basis. As found above, on a continuous basis the forward rate is $R_F = 11\%$. On an annually compounded basis, this is equivalent to 11.6278%, which is derived from the equation $e^{-.11} = (1 + R)^{-1}$.

Now, since we had agreed to get 12% when we first did the deal but now if we entered into an identical agreement to exchange cash flows at the same points in time we'd have to accept an 11.6278% rate, the value of the FRA is simply the notional amount times this difference times the length of the time period, or:

$$10{,}000{,}000(.12 - .116278)(T_2 - T_1)$$

but on a present value basis since the extra cash flow we are getting doesn't come until $T = 2$. To discount that, we use the current 2-year spot rate, 10%, and get the value of the FRA as:

$$10{,}000{,}000(.12 - .116278)(2 - 1)e^{-.1(2)} = \$30{,}473$$

Did you notice the subtle change in the definition of the time periods? The question said that the expiration date of the FRA was at $T = 2.5$ but we were valuing it 6 months later. Therefore, as of the point in time that we are valuing the FRA the time periods are each 6 months shorter.

Question 5. Assume you have a 2-year, 5% coupon bond with annual coupons and a \$1,000 face value. Determine the modified and Macaulay durations of the bond assuming that current yield is 5% on a continuously compounded basis.

Solution. The modified duration for a bond with continuous compounding is given by the formula,

$$D = \frac{\sum t_i c_i e^{-yt_i}}{B}$$

As shown below, the bond price is $B = 997.64$ and the duration is $D = 1{,}947.72/997.64 = 1.952$. Note that in the case of continuously compounded discount rates, the modified and Macaulay durations are identical.

Period	T	Cash Flow	PV(CF)	T*PV(CF)
1	1	50	47.56	47.56
2	2	1,050	950.08	1,900.16
			997.64	1,947.72
			Duration	**1.952**

Question 6. Calculate the convexity for the bond in the previous question.

Solution. The formula for convexity is as follows when you have continuous compounding:

$$C = \frac{\sum t_i^2 c_i e^{-yt_i}}{B}$$

Note that this formula is a little bit different with continuous compounding than it was with annual or semi-annual compounding. Recall that with annual compounding, the term t^2 was $t(t+1)$. In this case, the resulting calculation is as follows:

Period	T	Cash Flow	PV(CF)	$T^2PV(CF)$
1	1	50	47.56	47.56
2	2	1,050	950.08	3,800.32
			997.64	3,847.88
			Convexity	**3.857**

Recommended Textbook Problems

I strongly recommend working all of the end-of-chapter questions from the Hull textbook. But due to time constraints this may not be feasible. Therefore, at a minimum you should review all of the numerical examples that appear throughout the main text and the following end-of-chapter questions (especially the ones in bold):

1, **3, 5**, 6, **11**, 13, 14, 16, **17, 22, 23**

Hull Chapter 7: Swaps

Interest Rate Swap Mechanics

A swap is simply an agreement between two parties to exchange a series of cash flows at different points in time. A basic, plain vanilla, interest rate swap involves one party paying a fixed rate of interest on a notional principal amount to the other party and at the same time receiving from that party a variable rate of interest on the same notional principal amount.

The floating rate used in swaps could in theory be based on any reference rate, but in most swaps the London Interbank Offered Rate (LIBOR) is used. LIBOR is the rate at which banks are willing to lend to other highly rated banks for short term deposits of different maturities (one month, 3 month, 6 month, etc.).

In a typical swap, the principal amounts are not actually exchanged since they are the same amount for both parties, hence the term *notional principal*.

For the floating rate payer, the rate that they pay is set at the beginning of each period but the actual payment is made at the end of the period.

Example

Suppose Party A and Party B entered into a 1 year interest rate swap on a notional principal of $100 million, with A agreeing to make semi-annual payments at a fixed rate of 6% per annum and B agreeing to pay 6 month LIBOR. Assume further that the 6 month LIBOR rate is currently 5% and for the sake of this example that 6 month LIBOR six months from now will be 7%. Assume the rates are quoted on an annual basis assuming semi-annual compounding.

The cash flows for Party A and Party B are as shown in Table 1.

TABLE 1. Sample Swap Cash Flows

Party A				
Time	Rate	Pay	Receive	Net Cash Flow
6 Months	6.00%	3,000,000	2,500,000	-500,000
12 Months	6.00%	3,000,000	3,500,000	500,000
Party B				
Time	Rate	Pay	Receive	Net Cash Flow
6 Months	5.00%	2,500,000	3,000,000	500,000
12 Months	7.00%	3,500,000	3,000,000	-500,000

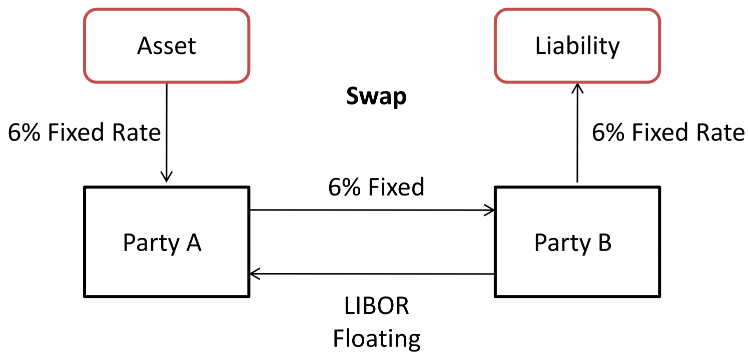
Uses of Interest Rate Swaps

Interest rate swaps play a critical role in interest rate risk management, allowing parties to very efficiently manage the interest rate sensitivity of their assets and/or liabilities.

Transforming a Liability

A swap like the one shown above can be used to help Party B transform an existing liability. For example, consider the situation depicted in Figure 1.

FIGURE 1. Swap Example



Assume that Party B had previously borrowed $100 million and agreed to pay the lender semi-annual interest payments at a rate of 6% per annum. If Party B would prefer to have floating rate payment obligations (for instance if their available cash flow fluctuated with interest rates), then they could enter into this swap with Party A. Since they receive from Party A each period the exact amount they need to pay their lender, they don't have to worry about whether their cash flow is sufficient to make the fixed rate payments. Instead, their payment obligation is simply to make the floating rate payments to Party A.

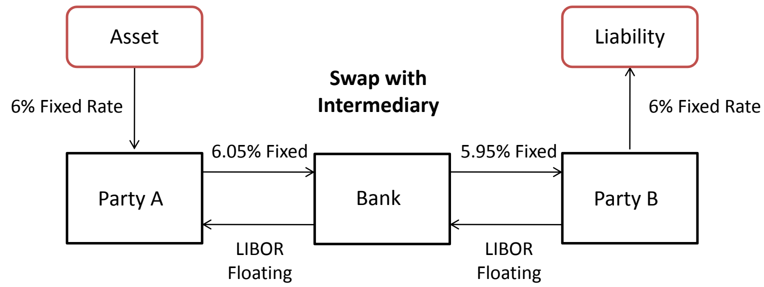
Transforming an Asset

The same arguments can be used to show that Party A in Figure 1 has essentially transformed an asset paying him a fixed rate of interest into a floating rate asset paying a variable rate of interest. If he owned a $100 million face value bond paying fixed 6% interest on a semi-annual basis, then this swap allows him to simply pay those cash flows to Party B and instead receive cash flows tied to LIBOR. This is shown in Figure 1.

Financial Intermediary

While any two parties are free to enter into swap agreements with each other, typically it is done with a financial intermediary (a swaps dealer) standing between the two of them. The dealer assumes the credit risk of each party and facilitates the netting of a huge number of offsetting transactions. The intermediary will typically charge a fee of a few basis points to each party. We can insert an intermediary into the previous example and assume that he charges each party 5 basis points, making Party A pay 6.05% fixed rate to receive LIBOR and paying Party B only 5.95% in exchange for LIBOR. This is shown in Figure 2 on the next page.

FIGURE 2. Swap Example with Intermediary



Swap Pricing

In a moment we will see how the fixed rates on swaps are determined, but note that swap dealers (market makers) will provide counterparties with pricing schedules which indicate the fixed rates they are willing to pay to receive LIBOR (bid rate) or the fixed rates they would like to receive to pay LIBOR (offer rate). The differences in these bid and offer rates will typically be in the range of a few basis points and reflects the fees noted earlier.

Swap Rate

The Swap Rate is the average of the bid and offer *fixed* rates on swaps that pay LIBOR as the floating rate.

The swap rate is not a risk-free rate, but it is close to risk-free because it generally reflects the credit risk of a highly-rated financial institution (AA, for example). Further, because the swap payments are successive payments from AA-rated financial institutions which are AA-rated at the start of each successive period, the credit risk of a 5-year swap is less than that of a 5-year loan to a financial institution that is AA-rated at the start but which may be lower rated in subsequent periods.

Swap Spread

The fixed rates are often quoted in relation to the U.S. Treasury rates for bonds with the same maturity, with the difference between the rates referred to as the *Swap Spread*. For instance, if the 3 year U.S. Treasury rate (the yield on a U.S. Treasury bond that matures in 3 years and sells at par) is 5% and the fixed rate for a swap is 5.25%, the swap rate would be quoted as UST+25 basis points.

Day Count Conventions

Recall that the rates for fixed income instruments typically follow particular, and often different, day count conventions regarding the number of days assumed to exist in a month or year. The day count conventions will be specified in the swap agreement and the actual payments will be adjusted each period to reflect this.

Business Day Conventions

The swap agreement will also indicate how to treat weekends and holidays (including which country's holidays to observe) with respect to when the cash flows are to be paid. For instance, it may specify that if the scheduled payment occurs on a Saturday, the payment is due the following business day. Or it may specify the preceding business day, etc.

Comparative Advantage Argument

One argument for the swap market to exist is the fact that different parties may have to pay comparatively different rates to borrow money at fixed rates or floating rates, so a swap between them may allow one or both parties to benefit from borrowing in the most efficient fashion and then entering into a swap with each other. To determine which market each party borrows in, we need to determine which market offers them a *comparative advantage*.

Numerical Example

Company A and Company B each have to borrow $10 million and each face the following borrowing costs in the fixed-rate and floating-rate loan markets:

TABLE 2. Borrowing Costs

	Company A	Company B
Fixed Rate	7.20%	9.20%
Floating Rate	LIBOR + .4%	LIBOR + 1.4%

Company A would prefer to borrow in the floating rate market and Company B would prefer to borrow in the fixed rate market. We want to determine the terms of a swap that they could enter into, without an intermediary, so that both companies are better off. We'll assume that whatever net savings are achieved are split equally between the two companies.

- Step 1: We start by checking to see whether there is a comparative advantage:

TABLE 3. Determining Comparative Advantage

	Company A	Company B	Difference
Fixed Rate	7.20%	9.20%	2.00%
Floating Rate	LIBOR + .4%	LIBOR + 1.4%	1.00%
Comparative Advantage			**1.00%**

A pays 2.00% less than B in the fixed rate market and only 1.00% less in the floating rate market. They pay "more less" in the fixed rate market so that is where they should borrow. If that's where they want to borrow then you are done. But here we are told that Company A wants a floating rate loan, which they can achieve by using a swap where they pay a floating rate (and receive a fixed rate).

- Step 2: Determine how the savings will be split.

  The comparative advantage calculation tells you what net savings is available for A and B to split. The net savings is just the difference between the 2.00% advantage A has in the fixed rate market and the 1.00% advantage A has in the floating rate market, for a 2.00% - 1.00% = 1.00% comparative advantage. This is the amount that is on the table to negotiate, but we will assume here they split it equally.

- Step 3a: Find the terms that Company A is willing to pay in the swap.

  If A borrows in the fixed rate market they have to pay 7.2%. They want to pay LIBOR plus x% (a floating rate) instead, so we know they want to RECEIVE the fixed rate in the swap and PAY the floating rate in the swap.

  We also know that we want the NET effect to be .5% cheaper costs than they would have to pay if they just borrowed in the floating rate market, so we know we want the net effect to be paying LIBOR + .4% − .5% = LIBOR − .1%.

  How can that be done? Simply by PAYING LIBOR - .1% in the swap and RECEIVING 7.2% in the swap. The 7.2% swap cash flow covers the 7.2% fixed rate they pay on the loan and the net effect is just the LIBOR-.1% payment.

- Step 3b: Find the terms that Company B is willing to pay in the swap.

  Company B borrows at L+1.4% in the floating rate market (that's where their comparative advantage is) but they really wanted to borrow at a fixed rate. We know that they have to get a net .5% savings (from Step 2), so we know they have to pay 9.2% - .5% = 8.7% on a net basis. They can do this by simply receiving LIBOR + 1.4% in the swap and paying 8.7% in the swap. The swap cash flows they receive offsets their borrowing costs and so the net payment for them is a fixed 8.7%.

  I showed this step for Company B, but obviously it wasn't really needed because, without an intermediary, Company B's swap cash flows are just the flip side of Company A's. They look different only because we haven't taken the (optional) step to simplify the terms so that the floating rate payment is zero.

- Step 4: Adjust terms so that floating payment is LIBOR.

  Here, instead of saying Company A receives 7.2% and pays LIBOR - .1% we could just say that they receive 7.3% and pay LIBOR. Similarly, Company B Receives LIBOR and pays 7.3%.

  This final adjustment is optional, but making the floating side equal to LIBOR "flat" makes it easier to value the swap, so it's good practice to do it.

Flaw in the Comparative Advantage Argument

The text notes that this argument is a bit flawed. By now the swap market is so big and so efficient that any of these arbitrage opportunities between markets ought to have gone away.

The reason for the apparent arbitrage is more related to the credit risk of the parties and the subtle difference between borrowing at a floating rate and at a fixed rate. Typically, in floating rate loans the lender will have the ability to reset the spread if the credit rating of the borrower changes, or in extreme cases even refuse to roll over the loan on the reset date. In contrast, lenders in fixed rate loans do not have the ability to change the rate during the term of the loan. So in the floating rate market, the lender is not really taking as much credit risk as would be the case in the fixed rate market.

Determining LIBOR/Swap Rates

As mentioned in Hull Chapter 4, LIBOR rates reflect rates that AA-rated banks charge each other for loans with maturities ranging from 1 to 12 months. For these maturities, the LIBOR rates can be observed. For longer maturities, LIBOR rates are generally determined using swap rates.

The *swap rates* are the fixed rates paid in a fixed-for-floating interest rate swap such that the value of the fixed rate and floating rate sides of the swap are both equal to the notional principal amount. This was described in Hull Chapter 4 as the par yield.

Using market swap rates, we can estimate a yield curve made up of LIBOR par yields and use this, just as we did before with U.S. Treasury par yields, to estimate a LIBOR zero coupon term structure.

Valuation of Interest Rate Swaps

There are two ways to determine the value of a swap that has already been entered into. One way is to think of a swap as the exchange of a fixed rate bond for a floating rate bond. The other way is to view a swap as a series of Forward Rate Agreements.

Value as an Exchange of Bonds

Although in a swap the notional principal is not typically exchanged, it would be irrelevant for valuation purposes even if it were since the amount is the same for both parties. But if you assume that it is exchanged, then the fixed rate side is like a fixed rate bond and the floating rate side is like a floating rate bond. Therefore, if B_{fl} is the value of the floating rate bond and B_{fix} is the value of a fixed rate bond, then the value of the swap is simply:

$$V_{swap} = B_{fl} - B_{fix}$$

for the person receiving the floating rate and paying the fixed rate.

We determine the value of the fixed rate bond the same as we always value coupon bonds, being careful to note that the Hull text uses continuous compounding when calculating present values.

For the floating rate bond, note that on the date the coupon rate is reset, the floating rate bond will always trade at par (recall a bond with coupons equal to the yield trades at par and on the reset date the coupon will always equal the yield). Once the rate is set, up until the next coupon reset date the floating rate payment is fixed at k^*, for instance. We also know that on the next reset date the bond will trade at par again and the value will again be L, where L is the notional principal of the swap. But the value at some point between the reset dates can fluctuate and is always worth:

$$\text{Value Between Resets} = (L + k^*)e^{-rt}$$

where L and k^* are the known values of the cash flows as of the next reset date, t is the time until the next reset date and r is the zero coupon rate for time t.

Note that since the payments have credit risk, you would typically use LIBOR zero coupon rates (rather than U.S. Treasury rates) to determine both the fixed and floating rate bond values.

Value as Series of FRAs

Another way to view the value of a swap is to value a series of FRAs.

Recall from Chapter 4 that the *value* of an FRA is found by valuing an agreement to exchange a payment based on a fixed rate of interest for a payment based on whatever the market rate turns out to be for the time period. To determine this value, we need to rely on the risk neutral approach whereby we assume that the forward rate will be realized. Since a swap actually is an exchange of a fixed rate for "whatever the rate turns out to be" (i.e. a floating rate) the valuation of a swap is just the valuation of a series of FRAs. Therefore, we can value each piece of the swap by assuming the forward rates are realized and calculate the present value of the net cash flows.

Note that in these situations, you will typically have to calculate the continuously compounded LIBOR forward rate first and then use that to determine the semi-annually compounded LIBOR forward rate for the purposes of determining the swap cash flows.

Overnight Indexed Swaps

An overnight indexed swap (OIS) is a swap where a fixed rate is exchanged for the geometric average overnight rate (e.g. the Fed Funds rate) during the term of the swap. The fixed rate paid in such a swap is referred to as the OIS rate.

Suppose a bank were to borrow in the overnight market for three months, rolling over the borrowing each day. They could use the money borrowed to lend to another party for three months at the rate of LIBOR and at the same time enter into an OIS to receive the overnight rate and pay a fixed rate (the OIS rate). The net effect will be to receive LIBOR for three months and pay the OIS rate. The difference in these two rates (the OIS rate is usually lower than LIBOR) will be the bank's compensation for assuming the credit risk of the bank they made the LIBOR loan to.

Why is this important? Because fluctuations in the difference between LIBOR and the OIS rate (the LIBOR-OIS spread) provide an indication of the credit risk of lending to major banks and is a signal of the degree of stress in the financial markets. This spread is typically 10 basis points or so, but during the 2007–2009 financial crisis is reached 364 basis points. After returning to normal levels it rose to over 30 basis points in June 2010 as a result of the European financial crisis that impacted Greece and other European countries.

The OIS rate is now regarded as a better proxy for the risk-free rate than LIBOR.

Currency Swaps

These are identical conceptually to interest rate swaps, with the obvious difference that the swap payments on each side of the transaction are made in different currencies. But all of the points made about interest rate swaps carry over directly to currency swaps — how they are used, how they are valued, their decomposition into a series of forward contracts, etc.

Credit Risk

The credit risk in a swap is the risk that the counterparty does not have the ability to make its payments to you. Since there will be scenarios where the net payments under a swap involve you paying the counterparty and scenarios where the net payments involve the counterparty paying you, the value of a swap at any time can be either positive or negative. It is only when the swap value is positive (i.e. when the net payments are owed to you) that there is credit exposure to the counterparty.

The credit crisis of 2007-2009 highlighted an important aspect of credit risk in swaps and other derivative securities — the threat that default by one counterparty can ripple through the financial system and cause multiple counterparties to default. One way to mitigate this risk is to move more swap and derivative transactions to clearing houses that act as intermediaries between parties, as opposed to direct transactions between counterparties. The clearing houses can establish collateral/margin requirements and monitor them continuously so that defaults don't ripple through the financial system and cause systemic risk.

Other Types of Swaps

You should be familiar with other types of swaps, including:

- Variations in Interest Rate Swaps
  - Different tenors

  - Alternative floating rates, such as the commercial paper rate
  - Floating for Floating
  - Amortizing Principal
  - Forward Swaps
  - Constant Maturity Swaps (CMS, CMT)
  - Compounding Swap
  - LIBOR in Arrears
  - Accrual Swap

- Variations in Currency Swaps
  - Cross Currency Interest Rate (Fixed for Floating in different currencies)
  - Floating for Floating
  - Quanto

- Equity Swaps

- Options
  - Extendable
  - Putable
  - Swaption

- Others
  - Commodity Swaps
  - Volatility Swaps

Practice Questions

Suppose you had previously entered into a 5 year fixed for floating interest rate swap where you agreed to pay a fixed annual rate of 5.6% on a notional balance of $10 million and receive 6-month LIBOR plus .4% on the same notional amount.

Assume the 6 month LIBOR rate was 5.1% on the day you entered into the swap, was 5.0% 6 months after inception and was 5.2% 12 months after inception.

What were the cash flows at inception, at 6 months after inception and at 12 months after inception? Ignore day-count conventions and assume that all rates given were bond equivalent annual rates. Inception: No cash flows occur.

6 Months after Inception: Cash flows are based on the fixed rate of 5.6% and the 6-month LIBOR rate of 5.1% at inception. That is, the floating rate payment is set 6 months before the payment is paid, using the floating rate at that time.

As the fixed rate payer, you will have paid 2.8%(10,000,000) = \$280,000 and you will have received (5.1% + .4%)/2(10,000,000) = \$275,000. The net payment would have been a \$5,000 payment from you to your counterparty.

12 Months after Inception: The payment at 12 months would have been based on the same fixed rate and the 6-month LIBOR rate 6 months earlier, or 5.0%. The net payment would have been \$280,000 − \$270,000 = \$10,000 from you to your counterparty.

Note that technically, day-count conventions are specified in the swap confirmation and both the fixed and floating rate payments are adjusted accordingly. Typically, the floating rate uses an Actual/360 convention and the fixed rate side uses Actual/365.

Question 1. Suppose that you are the CFO of an industrial company, ABC, looking to borrow at floating rates of interest that will be correlated with the cyclicality of your business. You can borrow at fixed rates of 7.5% or at floating rates of LIBOR plus 20 basis points. Another company, XYZ, with a lower credit rating prefers not to introduce uncertainty with respect to their interest obligations and would prefer to borrow at fixed rates. They can borrow at fixed rates of 9% or at floating rates of LIBOR plus 100 basis points. Determine a swap that could make each of you better off.

Solution. Step 1: Determine if there is a comparative advantage.

Note that while ABC can borrow at lower rates than XYZ in both the fixed and the floating rate markets, there is a 150 basis point spread between their respective rates in the fixed rate market and only an 80 basis points in the floating rate market. This gives ABC a 70 basis point comparative advantage:

TABLE 4. Determining Comparative Advantage

	ABC	XYZ	ABC's Absolute Advantage
Fixed Rate	7.50%	9.00%	1.50%
Floating Rate	LIBOR + .2%	LIBOR + 1%	0.80%
Comparative Advantage			**0.70%**

Notice that ABC can pay lower rates in either market, however they pay “more less” in the fixed rate market. This gives them a comparative advantage in the fixed rate market, so this is where they should borrow.

Step 2: Determine how the savings should be split.

Using the swap market, ABC can borrow in the fixed rate market, where they have a comparative advantage and then enter into a swap with XYZ to convert their payments to floating rate payments. How the 70 basis points of potential savings will be split between the two companies is a matter for negotiation, however any split will make both of them better off.

Suppose for simplicity that they split the 70 basis point gain equally and each gets 35 basis points.

Step 3a: Determine the terms that ABC is willing to pay.

ABC borrows in the market where they have the comparative advantage (the fixed rate market) at 7.5% and agrees to pay LIBOR plus some spread to XYZ in exchange for XYZ's fixed rate payments. This will leave ABC paying fixed rates to borrow, receiving fixed rates from the swap and paying floating rates in the swap. Overall, ABC will be a net floating rate payer.

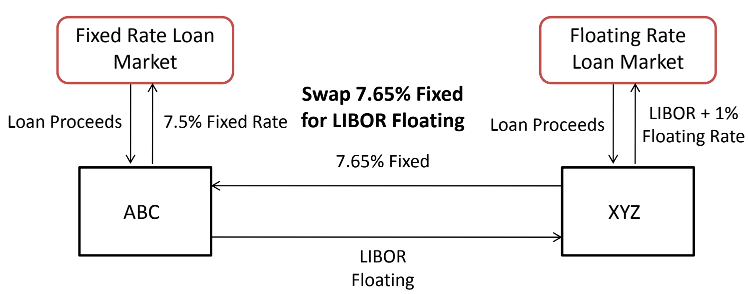
Had they borrowed in the floating rate market to begin with, they would have had to pay LIBOR + .2%. If they are going to save 35 basis points (their share of the total potential savings), they will have to pay LIBOR - .15% overall, which they can do if they enter a swap in which they receive 7.5% (the amount they they have to pay on their borrowing) and pay LIBOR - .15% (35 basis points lower than what they would have paid in the floating rate market).

Step 3b: Determine the terms that XYZ is willing to pay.

Here, we can do the same thing as we did for ABC, but we don't really have to since there is no intermediary and therefore we already know the terms of the swap — they are the flip side of the terms ABC is willing to pay: XYZ pays 7.5% and receives LIBOR - .15%. Since they pay LIBOR+1% in the floating rate market to borrow, their net payments are LIBOR+1% +7.5% - (LIBOR - .15%) = 8.65%, which is 35 basis points below what they could have done on their own (9%).

Step 4: As an optional last step, you could restate the terms of the swap so that one side pays LIBOR flat (with no spread), in which case the terms would be Pay 7.65% and Receive LIBOR.

FIGURE 3. Key Components of Net Cash Flows



Question 2. Company A and Company B each have to borrow \$10 million and each face the following borrowing costs in the fixed-rate and floating-rate loan markets:

TABLE 5. Borrowing Costs

	Company A	Company B
Fixed Rate	7.20%	9.20%
Floating Rate	LIBOR + .4%	LIBOR + 1.4%

Company A would prefer to borrow in the floating rate market and Company B would prefer to borrow in the fixed rate market. Determine the terms of a swap that they could enter into, through Bank X as intermediary, so that both companies are better off and the bank earns a .1% fee on each swap it enters into. If there are net savings to be achieved, assume that Company A gets 75% of it and Company B gets 25% of it.

Solution. Step 1: Determine if there is a comparative advantage.

TABLE 6. Determining Comparative Advantage

	Company A	Company B	Difference
Fixed Rate	7.20%	9.20%	2.00%
Floating Rate	LIBOR + .4%	LIBOR + 1.4%	1.00%
Comparative Advantage			**1.00%**

A pays 2.00% less than B in the fixed rate market and only 1.00% less in the floating rate market. They pay “more less” in the fixed rate market so that is where they should borrow. If that’s where they want to borrow then you are done. But here we are told that Company A wants a floating rate loan. That’s why we know they want to use a swap.

Step 2: Determine how the savings will be split. The comparative advantage calculation tells you what net savings is available for the three parties to split. The net savings (ignoring the bank for a moment) is just the difference between the 2.00% advantage A has in the fixed rate market and the 1.00% advantage A has in the floating rate market, for a 2.00% - 1.00% = 1.00% comparative advantage. This is the amount that is on the table to negotiate.

We are told the bank intermediary wants to make .2% (between the two swaps), so that tells us that there is .8% savings for A and B to split. Often the questions will assume they split it equally, but here I said to assume that they split it 75%/25%, so A will get .6% and B will get .2%.

Step 3a: Find the terms that each party pays in the swap, starting with Company A. If A borrows in the fixed rate market they have to pay 7.2%. They want to pay LIBOR plus x% (a floating rate) instead, so we know they want to RECEIVE the fixed rate in the swap and PAY the floating rate in the swap. We also know that we want the NET effect to be .6% cheaper costs than they would have to pay if they just borrowed in the floating rate market, so we know we want the net effect to be paying LIBOR + .4% − .6% = LIBOR − .2%.

How can that be done? Simply by PAYING LIBOR - .2% in the swap and RECEIVING 7.2% in the swap. The 7.2% swap cash flow covers the 7.2% fixed rate they pay on the loan and the net effect is just the LIBOR-.2% payment.

Step 3b is the identical thing, but this time for B. They borrow at L+.1.4% in the floating rate market (that's where their comparative advantage is) but they really wanted to borrow at a fixed rate. We know that they have to get a net .2% savings (from Step 2), so we know they have to pay 9.2% - .2% = 9.0% on a net basis. They can do this by simply receiving LIBOR + 1.4% in the swap and paying 9% in the swap. The swap cash flows they receive offsets their borrowing costs and so the net payment for them is a fixed 9%.

Step 4: Notice that this is ALL there is to it. We are really done. But it is safe to check to make sure everything is in balance. We can do this by confirming what happens at the intermediary. From their swap with A they receive LIBOR - .2% and pay 7.2%. From their swap with B they pay LIBOR+1.4% and receive 9%. That means on a net basis they pay LIBOR+8.6% and receive LIBOR+8.8%. The LIBOR they pay and receive cancel out and the net effect is that they receive 8.8% − 8.6% = .2% fixed. That's exactly what we wanted to happen.

We're done because everyone is happy. A pays .6% less than if they borrowed directly in the floating rate market. B pays .2% less than if they borrowed directly in the fixed rate market. And the bank receives a fixed .2% fee (assuming neither party defaults on them).

Step 5: One last optional step is to convert the swap terms so that the floating payment is just LIBOR. We can add or subtract any fixed amount to both sides of a swap and not change the economics at all. So, for instance, paying 8.3% and getting LIBOR +.3% is the same as paying 9.3% and getting LIBOR+1.3%. It is also the same as paying 8% and getting LIBOR.

In this case, we could adjust the swap terms as follows:

- A's Swap with Bank: Pay LIBOR - .2%/Receive 7.2% ⇒ Pay LIBOR/Receive 7.4%
- B's Swap with Bank: Pay 9%/Receive LIBOR + 1.4% ⇒ Pay 7.6%/Receive LIBOR%

This final adjustment is optional, but making the floating side equal to LIBOR "flat" makes it easier to value the swap, so it's good practice to do it.

Question 3. Three months ago you entered into an interest rate swap whereby you agreed to pay a fixed rate of 4% with semi-annual compounding and receive floating rate payments equal to 6-month LIBOR, also with semi-annual compounding. The notional principal of the swap was $10 million, the maturity of the swap was 12 months, with the next payment dates in 3 months and 9 months. At the inception of the swap, the 6-month LIBOR rate was 4.4% with semi-annual compounding and currently the continuously compounded LIBOR zero rates for 3 months and 9 months are 5% and 5.5%, respectively. Determine the current value of the swap in terms of the values of fixed and floating rate bonds, ignoring any complications resulting from day count or business day conventions.

Solution. To value the swap, you can view this as a long position in a floating rate bond and a short position in a fixed rate bond. For the fixed rate bond, the coupons are 4% and the principal is \$10 million and so the cash flows are simply 4%/2(10,000,000) = 200,000 in 3 months and 10,200,000 in 9 months. The value of this is found by discounting at the continuous LIBOR rates:

$$\begin{aligned} B_{fix} &= 200{,}000e^{-.05(.25)} + 10{,}200{,}000e^{-.055(.75)} \\ &= 9{,}985{,}325 \end{aligned}$$

For the floating side, the rate for the next payment has already been set at 4.4%/2 = 2.2% and the cash flow is simply 2.2%(10,000,000) = 220,000. We also know that on the next reset date, the floating rate bond will be worth par, or \$10 million. Therefore, the value today is simply:

$$\begin{aligned} B_{fl} &= 10{,}220{,}000e^{-.05(.25)} \\ &= 10{,}093{,}045 \end{aligned}$$

The resulting swap value is then $B_{fl} - B_{fix} = 107{,}720$.

Question 4. Determine the value of the swap in the previous question by interpreting the swap as a series of Forward Rate Agreements (FRA).

Solution. The value of an FRA is found by valuing an agreement to exchange a payment based on a fixed rate of interest for a payment based on whatever the market rate turns out to be for the time period. To determine this value, we need to rely on the risk neutral approach whereby we assume that the forward rate will be realized. Since a swap actually is an exchange of a fixed rate for "whatever the rate turns out to be" (i.e. a floating rate) the valuation of a swap is just the valuation of a series of FRAs. Therefore, we can value each piece of the swap by assuming the forward rates are realized and calculate the present value of the net cash flows.

For the first payment under the swap, we know that we will pay a fixed rate and receive the floating rate that was established at inception (4.4% on a bond equivalent basis). Therefore we know that on the first payment date the net payment to us will be 220,000 − 200,000 = 20,000. The present value of this is $20{,}000e^{-.05(.25)} = 19{,}752$.

For the second payment, we first need to determine the LIBOR forward rate. Using the continuously compounded LIBOR rates given for the 3 month and 9 month periods, we know that if F is the forward rate then $e^{-.05(.25)}e^{-F(.5)} = e^{-.055(.75)}$ and from that we can get the formula we saw in Chapter 4,

$$F = \frac{R_2T_2 - R_1T_1}{T_2 - T_1} = \frac{.055(.75) - .05(.25)}{.75 - .25} = 5.75\%$$

But note that this is a continuously compounded rate and the swap payments are made using the semi-annual compounded rate. So we need the semi-annually compounded forward LIBOR rate, which we can find easily by noting that $e^{.0575(.5)} = 1 + (r/2)$ where r is the semi-annually compounded forward rate, or $r = 2(e^{.0575(.5)} - 1) = 5.833\%$.

Then, the second payment of the swap is simply the difference in the cash flows received and paid, assuming that the forward rate is realized. The present value of this is then simply:

$$PV(CF_2) = \left(\frac{5.833\%}{2}10{,}000{,}000 - \frac{4\%}{2}10{,}000{,}000\right) e^{-.055(.75)} = 87{,}968$$

The total swap value is then 19,752 + 87,968 = 107,720.

Notice that the formula for the present value of the second cash flow looks a little different from the formula we used in Hull Chapter 4. Using that formula, we would have written this as:

$$PV(CF_2) = L(R_K - R_F)(T_2 - T_1)e^{-R_2 T_2}$$

Aside from the change in the sign to denote that we are receiving the floating rate the only other thing I did differently was to combine annual rates (e.g. 5.833% and 4%) with the time period over which those rates are earned ($T_2 - T_1$). The way I did it is a bit more intuitive, but the formula is the same so you should not get confused.

Question 5. Three months ago you entered into an interest rate swap whereby you agreed to receive a fixed rate of 4.2% with semi-annual compounding and pay floating rate payments equal to 6-month LIBOR + .2%, also with semi-annual compounding. The notional principal of the swap was \$10 million, the maturity of the swap was 12 months. Since three months have passed, the next payment dates are in 3 months and 9 months. At the inception of the swap, the 6-month LIBOR rate was 4.4% with semi-annual compounding and currently the continuously compounded LIBOR zero rates for 3 months and 9 months are 5% and 5.5%, respectively. Determine the current value of the swap in terms of the values of fixed and floating rate bonds, ignoring any complications resulting from day count or business day conventions.

Solution. Note that this is really the same as the previous questions, with two minor differences. First, we are now paying floating and receiving fixed. Second, the rates paid are slightly different.

The key thing to notice though is that because the floating rate pays LIBOR plus .2% instead of just LIBOR flat, the value of the floating rate bond is not equal to par on the reset dates. Therefore, the best thing to do is to first adjust the payments so that one side pays LIBOR flat. Here, note that paying LIBOR plus .2% and receiving 4.2% is really just the same as paying LIBOR and receiving 4%. Then, the calculations are easier — especially because now this is the identical swap we valued in the previous question.

To value the swap, you can view this as a short position in a floating rate bond and a long position in a fixed rate bond. For the fixed rate bond, the coupons are 4% and the principal is \$10 million and so the cash flows are simply 4%/2(\$10,000,000) = \$200,000 in 3 months and \$10,200,000 in 9 months. The value of this is found by discounting at the continuous LIBOR rates,

$$B_{fix} = \$200{,}000e^{-.05(.25)} + \$10{,}200{,}000e^{-.055(.75)} = \$9{,}985{,}325$$

For the floating side, the rate for the next payment has already been set at $4.4\%/2 = 2.2\%$ and the cash flow is simply $2.2\%(\$10{,}000{,}000) = \$220{,}000$. We also know that on the next reset date, the floating rate bond will be worth par, or \$10 million. Therefore, the value today is simply,

$$B_{fl} = \$10{,}220{,}000e^{-.05(.25)} = \$10{,}093{,}045.$$

The resulting swap value is then $B_{fix} - B_{fl} = -\$107{,}720$.

Question 6. Suppose that you entered into a currency swap with 12 more months remaining in which you receive 6% semi-annually in dollars on a notional principal of \$10 million and pay 3% semi-annually in yen on a notional principal of 1,300 million yen. Assume that the continuously compounded zero rates in Japan are 1% and in the US are 2% for both 6 month and 12 month maturities. Also assume that the current exchange rate is 120 yen per dollar. Determine the value of the swap as the differences between the value of a dollar bond and yen bond.

Solution. Similar to what was done for interest rate swaps, we can view this swap as being a long position in a dollar bond and a short position in a yen bond. The dollar bond has cash flows in 6 months and 12 months equal to $(6\%/2)(\$10{,}000{,}000) = \$300{,}000$ and \$10,300,000. The present value of these amounts are found using the dollar zero rates:

$$B_{\$} = \$300{,}000e^{-.02(.5)} + \$10{,}300{,}000e^{-.02(1)} = \$10{,}393{,}061$$

For the yen bond, the cash flows are 19.5 million yen and 1,319.5 million yen and the present value of these at the yen interest rates is 1,325,773,499 yen. In dollars, converted at the current spot exchange rate of 120 yen per dollar, this is worth \$11,048,112 dollars.

The total value of the swap then is $-\$655{,}051$.

Question 7. Determine the value of the swap in the previous question by valuing the swap as a series of forward contracts.

As a reminder, the formula for the forward exchange rate is $F = Se^{(r-r_f)T}$ where S is the current exchange rate in dollars per foreign currency, r is the U.S. risk-free rate, r_f is the foreign risk-free rate and t is the expiration date.

Solution. To value the swap as a series of forward contracts, we first need the forward exchange rates for 6 month and 12 month maturities. Here, the spot exchange rate was given as 120 yen per dollar, which means that $S = 1/120 = .00833$ dollars per yen. The two forward exchange rates are:

$$F_{.5} = (.00833)e^{(.02-.01).5} = .0083751$$

$$F_1 = (.00833)e^{(.02-.01)1} = .00841708$$

If we assume that these forward rates are realized, we can see that the first exchange of payments, in dollars, is:

$$\$300{,}000 - 19{,}500{,}000(.0083751) = \$136{,}685$$

and the second exchange in dollars would be:

$$\$10{,}300{,}000 - \$1{,}319{,}500{,}000(.00841708) = -\$806{,}343$$

The present value of these cash flows, discounted using the dollar zero rates, is −\$655,051, just as before.

Recommended Textbook Problems

I strongly recommend working all of the end-of-chapter questions from the Hull textbook. But due to time constraints this may not be feasible. Therefore, at a minimum you should review all of the numerical examples that appear throughout the main text and the following end-of-chapter questions (especially the ones in bold):

1, 2, **3, 5**, 7, **16**

BKM Chapter 16: Managing Bond Portfolios

These notes reflect only Sections 16.3 — 16.4 of the BKM text. See the earlier notes for Sections 16.1 and 16.2 for important background explanations of duration and convexity that are needed to understand the material shown here.

Passive Investment Strategies

Passive investment strategies, as compared to active strategies, do not attempt to identify underpriced or overpriced bonds, but rather they assume that prices are fairly set. There are two types of passive bond investment strategies — Indexing and Immunization.

Indexing

The goal of indexing is to simply mirror the overall results of the broad market. The text describes three common indices, all of which contain a broad, market-value-weighted portfolio of bonds from various markets (government, corporate, mortgage backed, Yankee bonds issued by foreign corporations).

Unlike stock indices, the composition of bond indices change quite frequently as bonds mature, bonds are called, bonds default, etc. As a result, matching the index is more difficult and necessarily contains tracking error (deviations of the portfolio's return from the index return).

To create a passive portfolio that tracks one of these indices, it is necessary to break the index into broad maturity and sector categories and to form a portfolio with a similar maturity and sector profile as the index.

Immunization

This approach attempts to eliminate interest rate risk from a portfolio by either ensuring that changes in interest rates do not affect the total market value of the firm (duration based) or that changes in rates do not affect the expected future value (with reinvestment) of the portfolio.

Net Worth Immunization

This is covered in detail in the Noris and Feldblum readings. The point here is that the value of a firm's assets and the value of their liabilities both change as interest rates change. Therefore, to protect the firm against changes in net worth (their ability to meet future obligations) they need to ensure that the value of their assets and the value of their liabilities change by the same dollar amount as rates change. Since the sensitivity of asset and liability values to interest rate changes is measured by duration, this simply means that they need to set the dollar duration of their assets (the duration of the assets multiplied by the dollar value of the assets) equal to the dollar duration of their liabilities (the duration of the liabilities multiplied by the dollar value of the liabilities).

Target Date or Holding Period Immunization

The future value of a bond (or bond portfolio) on any future date depends on both the price of the bond at that point plus the future value of all the reinvested coupons. As interest rates rise, the value of the first portion falls (this is price risk) but the value of the second portion rises (this is reinvestment risk). If we are concerned about the total future value though, then we should ensure that these two effects cancel each other out.

It turns out that the two effects are exactly offsetting when the Macaulay duration of the bond equals the time period you are targeting. So if the bond's Macaulay duration equals D, then the value of the bond at time period $T = D$ is unaffected by small changes in the interest rate. Note that when the text refers to *duration* in this chapter, they are usually referring to Macaulay duration, not the modified duration.

Rebalancing

Immunization (when focused only on matching durations) works only for small changes in rates because of the effect of convexity. Another way to think of this is that as the rates change, so do the durations, meaning that once the rates change a bit, the durations are no longer equal and therefore the portfolios need to be *rebalanced*. So even if your intent is to simply match durations, this doesn't make interest rate risk management simple because durations change continuously.

Cash Flow Matching and Dedication

If immunization is so inexact and rebalancing is such a pain, why not just match cash flows — buy bonds with payout amounts and timing exactly equal to your payment obligations? For instance, you could follow *cash flow matching* and just buy a zero coupon bond that matures at the same time as the projected cash outlay. Or, in a multi-period context, use a *dedication strategy* that matches the cash flows on either zero coupon or coupon bonds to a whole series of obligations.

This would eliminate interest rate risk, but it would severely limit portfolio choice. Worse, if liability durations are very long you may not be able to find any assets with matching cash flows.

Other Problems with Conventional Immunization

Besides the fact that convexity causes duration to be an imprecise estimate of the change in value for large interest rate movements, you also need to note that what we've done so far assumes that the yields for each time period change by the same amount. This is equivalent to assuming a parallel shift in the term structure. Second, we ignored inflation and assumed that we wanted to lock in the nominal future value of our portfolio.

Active Bond Management

The text discusses two forms of active bond portfolio management, that attempt to improve returns either by forecasting future rates and buying assets that will respond favorably to anticipated rate changes or by identifying mispriced bonds.

Bond Swaps

Replacing one bond with another in a portfolio is known as a bond swap. There are five types discussed to take advantage of different views of what might happen to bond values:

- Substitution Swap — This involves replacing one bond with a nearly identical substitute with essentially the same coupons, maturity, credit quality, etc. but with a lower price.

- Intermarket Spread Swap — This is done in anticipation of changes in spreads between two different sectors of the bond market. For instance, if the spread between corporate and government bonds is expected to narrow, then an investor might want to swap government bonds for corporate bonds.

- Rate Anticipation Swap — This is done in anticipation of changes in interest rates that might affect some bonds more than others. For instance, if you expect rates to decline, you might want a longer duration portfolio.

- Pure Yield Pickup Swap — This is done to replace low yielding bonds with higher yield bonds, not because of perceived mispricing but because of the desire to earn the higher yield (even if that means taking on more risk). Replacing short term government bonds with longer term government bonds when the yield curve is upward sloping is an example of this.

- Tax Swap — These are mentioned briefly and involve trying to achieve a tax benefit, for instance from the differences in the taxation of interest income versus capital gains (or price appreciation).

Note that the substitution and intermarket spread swaps are both done when you think that certain bonds are temporarily mispriced relative to other bonds; the rate anticipation swap and the yield pickup swap are done when you are trying to profit from your views on interest rates.

To see why interest rate views would lead to a rate anticipation swap, note that the shorter the duration of a bond, the earlier on average are the cash flows. When would you want cash flows early? When you think that you can reinvest those at very high rates. Therefore, if you think rates will rise, you might want to swap your long duration bonds for short duration bonds. Similarly, we know that long duration bonds have more risk than short duration bonds because they are more sensitive to interest rate movements. But if you think that rates are not going to change much, you might be willing to take on this added risk by swapping a short duration low yield bond for a long duration high yield bond, a yield pickup swap, thus getting paid a higher yield for something that you perceive to be of little or no extra risk.

Horizon Analysis

When an investor has a short investment horizon, they may choose to invest in longer term bonds and forecast what they think the yield curve will look like at the end of their horizon. Suppose the yield curve originally slopes upward and doesn't change over time. An investor purchasing a 20-year bond with the intent to sell it in 2 years would expect to not only earn the coupon interest and reinvest it over the two year period, but also receive a capital gain from the change in the bond price during the two year period. The bond price change arises from two sources. First, the future cash flows will be discounted for a shorter time period because the original 20 year bond will now mature in 18 years (at the horizon date). Second, the discount rate will be lower because of the originally upward sloping yield curve (this latter part gives rise to the term *riding the yield curve*).

Recall that one reason for an upward sloping yield curve, according to the expectations hypothesis, is the fact that rates are in fact expected to rise. If they do rise, then riding the yield curve will not necessarily result in a higher holding period return. This is just a reminder that investing long term may result in higher yields, but this is often at the expense of higher risk of loss due to changes in the yield curve.

Practice Questions

Question 1. Assume your firm has a bond portfolio containing a single bond with a \$10,000 face value, 8% coupons paid annually and a 5 year maturity. Also assume that you have a liability of a single payment of \$13,936 to make in exactly 4.312 years. Assume that the effective yield is also 8% on an annual basis — i.e. assume the present value for a payment in 1 year is 1.08^{-1} not 1.04^{-2}. What is the net worth of the firm? How much does the net worth of this firm change if interest rates suddenly increase to 8.1%? Assume the term structure is flat.

Solution. Notice the assumption is that the coupons are paid annually, so if interest rates are currently 8%, then it is easy to show that the present value of the bond is:

$$B = \frac{800}{1.08^1} + \frac{800}{1.08^2} + \frac{800}{1.08^3} + \frac{800}{1.08^4} + \frac{10{,}800}{1.08^5} = \$10{,}000$$

The present value of the liability is:

$$\frac{13{,}936}{1.08^{4.312}} = \$10{,}000$$

Therefore the net worth of this firm is \$10,000 − \$10,000 = 0.

The flat term structure tells us that all interest rates for all time periods change to 8.1%. We simply recalculate the present value of the bond at the new yield of 8.1% to get $B = \$9{,}960$.

Similarly, the present value of the liability changes to:

$$\frac{13{,}936}{1.081^{4.312}} = \$9{,}960$$

Therefore, the net worth remains at 0.

Question 2. Explain your results from the previous question in terms of duration.

Solution. If you calculate the Macaulay duration of the bond, you will see that it is equal to 4.312. Also note that the Macaulay duration for a zero coupon bond with a single cash flow at time t is equal to t. Therefore, the Macaulay duration of the liability is equal to 4.312 also.

Since the durations are equal, and the present value of the asset and the present value of the liability are equal, then the values of both change by the same amounts when rates change, thus the net worth is *immunized*.

Question 3. Assume you own a bond with $1,000 in face value, 11% coupons paid semi-annually and 5 years to maturity. Interest rates are currently 10% and the term structure is flat. What is the total value of this bond, including reinvestment of coupons, two years from today if rates are unchanged? What if rates immediately increase to 11%?

Solution. The future value will include the future value of the coupons plus the value of the bond two years from today. The first part is found by noting that we will receive coupons of $55 in 6, 12, 18 and 24 months. These will have future values at $T = 24$ months of:

$$\begin{aligned}\text{Future Value of Coupons} &= 55(1.05^3) + 55(1.05^2) + 55(1.05^1) + 55(1.05^0)\\ &= \$237.06\end{aligned}$$

Note that to get the future values, the coupon we receive at $T = 24$ months does not earn any interest, the one before that earns 6 months of interest, the one before that earns 12 months of interest, etc. For the value of the bond in two years, note that this is now just a 3-year bond with an 11% coupon, so its value is easily found to be $1,025.4. The total future value is $1,262.4.

If rates increase to 11%, we simply redo those calculations, noting that the reinvestment of coupons is now done at 11% and the present value of the bond at 2 years is now found by discounting the remaining cash flows at 11%. These two values are easy to show to equal $238.8 and $1,000 for a total of $1,238.8. Note that the value of the bond at $T = 24$ months is easy to calculate because the coupon and the yield are the same, so its value is the face value.

So if rates rise, we earn slightly more from reinvesting the coupons at the higher rate, but we lose a bit more than that due to the drop in value of the bond.

Question 4. Redo the previous question but find the values at $t = 4$ years. Note that the Macaulay duration of this bond could be found to equal 4.0.

Solution. Here, we just have to redo the same calculations. If the rates stay at 10%, the coupons have a future value equal to:

$$\text{Future Value of Coupons} = \sum_{t=1}^{8} 55(1.05^{8-t}) = \$525.2$$

The present value of the bond at that point is simply the present value of the remaining cash flows, or:

$$B = \frac{55}{1.05} + \frac{1{,}055}{1.05^2} = \$1{,}009.3$$

The total is then \$1,534.5.

Then, do the same thing again with rates equal to 11%. The coupons have a future value of \$534.7 and the bond has a value at that point of \$1,000 (equal to the face value). The total value is \$1,534.7.

In this case, the total value in 4 years is unchanged when the rates change because the increased income from reinvesting coupons at higher rates completely offsets the loss in value of the bond. This occurs only at the point in time equal to the Macaulay Duration, which in this case was equal to 4.0.

Question 5. Suppose you are the CFO of a Savings & Loan with initial capitalization of \$200 million. You raise \$500 million from depositors, agreeing to pay them a floating rate of interest each period but allowing them to withdraw funds at any time.

You then use the \$500 million to invest in mortgages with interest rates of 6% annually (i.e. providing the funds up front to homeowners in exchange for their promises to pay principal and interest monthly for 30 years).

Overall then, you will have \$700 million in assets, \$200 million of which we'll assume is deposited in short term risk free securities paying floating rates of interest and \$500 million of which is in mortgages. In addition, you have \$500 million in liabilities, also paying floating rates of interest.

Determine the duration of the assets and liabilities assuming that mortgage rates are currently 6% and estimate what would happen to the S&L's net worth if mortgage rates suddenly increased to 10%.

Solution. The duration of the short-term assets will be approximately zero, since changes in rates are immediately reflected in the interest that is earned.

The duration of the mortgages can be found relatively easily. First, determine the total monthly cash flows assuming a present value of \$500 million, a fixed payment for 360 periods, no final payment and a 0.5% monthly rate of interest. We know that fixed payments for 360 months is just an annuity, so we can set the present value of this equal to \$500 million and solve for the monthly payment (PMT):

$$\begin{aligned}
\$500{,}000{,}000 &= \text{PMT}\,\frac{1}{r}\left[1-(1+r)^{-N}\right]\\
&= \text{PMT}\,\frac{1}{.005}\left[1-(1.005)^{-360}\right]
\end{aligned}$$

The monthly payment turns out to be \$2,997,753 for all mortgages combined.

Note that you could use your calculators to find this. Just enter the following values and then compute PMT:

$$N = 360, i = .5\%, PV = -500{,}000{,}000, FV = 0$$

Now, we need to calculate the duration of this stream of payments, which can be found to equal 10.72. To do this, I used the shortcut formula and found the present value of the annuity with payments of \$2,997,753 for 360 months when rates rose by .1% and when rates fell by .1%. I then found the duration as:

$$\begin{aligned}
D &\approx \frac{P_- - P_+}{2P(\Delta y)}\\
&= \frac{505{,}406{,}865 - 494{,}682{,}908}{2(500{,}000{,}000)(.1\%)}\\
&= 10.72
\end{aligned}$$

Therefore, the asset portfolio duration is the weighted average duration, or:

$$\text{Asset Portfolio Duration} = \frac{200}{700}(0) + \frac{500}{700}(10.72) = 7.66$$

Although the asset duration is 7.66, the liability duration is approximately zero because it pays a variable rate of interest and can be immediately withdrawn, suggesting a significant Asset-Liability duration mismatch.

If rates were to go immediately to 10%, then the liabilities would still be worth \$500 million (depositors would now earn 10% interest and could still remove the funds at any time). The floating rate assets in the asset portfolio would be worth \$200 million, but the mortgage values would drop to \$341.6 million, a drop of 32%. Note that based on the duration we would have estimated a drop of $(10.72)(.04) = 43\%$ without the convexity adjustment and 28.5% with the convexity adjustment. But of course, this problem assumes a substantial increase in the rates, so even with the convexity adjustment the estimates would be off a bit.

Nonetheless, note that the change in rates caused the net worth to decline from \$200 million to $\$200 + \$341 - \$500 = \$41$ million, nearly wiping out the entire S&L.

Note that this scenario is what happened to the S&L industry in the 1980's. S&L's had short term, low duration floating rate liabilities and long term, fixed rate high duration mortgage assets. When rates increased sharply in the early 1980's, the entire industry was hurt.

It is interesting to note that when they went to Congress for a bailout, they were told to "invest their way out of the problem" and were encouraged to invest in much riskier but higher expected return assets. One class that was particularly attractive was junk bonds, which offered very high promised yields but with very high credit risk such that the expected yields were probably no different than less risky assets. But their timing could not have been worse. When Mike Milken and his firm, Drexel Burnham, were indicted for alleged securities law violations, the S&L's and insurers who owned these assets were forced to liquidate their junk bond portfolios at fire sale prices, which led to even steeper losses for the S&L's and in the end a major crisis that ultimately required a government bailout.

Question 6. Suppose the S&L mentioned above realized their Asset-Liability duration mismatch was exposing them to excessive interest rate risk and wanted to manage this risk. One thing to do would be to sell some of the mortgage assets to someone. But that could be difficult to do and may force a recognition of a market value loss, so assuming they did not want to sell mortgages, how else could they resolve their Asset-Liability duration mismatch?

Solution. One method discussed in the text is to use swaps, whereby you agree to pay someone the fixed rate coupons you receive and in return get floating rate coupons from them. This effectively converts a fixed rate asset into a floating rate asset, without requiring any up-front cash outlay (usually) or any adjustment to the assets in the portfolio.

Another approach would be to invest in very high *negative* duration assets. For instance, Stripped Mortgage Interest Only (IO) securities have a large negative duration. When rates rise, refinancings and other sources of prepayment generally decline so the IO holders receive substantially more interest cash flows than originally expected. This causes the IO to increase in value as rates rise — the opposite affect than what usually occurs for bonds.

It is interesting to note that many S&L's actually did follow this strategy of using complex derivatives, especially CMOs and stripped mortgage securities like IOs and POs, to rebalance their portfolio durations. But as the crisis unfolded, many S&L executives looked reckless investing in such complex securities and some of them were severely penalized for their alleged recklessness (some even went to jail, though arguably that was due to other misdeeds).

Question 7. Suppose the term structure is upward sloping with one-year annual zero coupon yields equal to 5% and two-year annual zero coupon yields equal to 6%. Compare the expected one year holding period returns from investing in either a one-year zero coupon bond or a two-year zero coupon bond, each with a face value of $1,000, if the term structure is unchanged at the end of the year. What happens if we assume that the expectations hypothesis holds?

Solution. The prices of the one-year and two-year zero coupon bonds will be as follows:

$$B_1 = \frac{\$1{,}000}{1.05} = \$952.4$$

$$B_2 = \frac{\$1{,}000}{1.06^2} = \$890$$

If we invest in the one-year zero, then we will pay $952.4 and receive $1,000 in one year, resulting in a 5% holding period return. If we buy a two-year zero, we will pay $890 and then at the end of one year the bond will have one more year until maturity and will have a price of:

$$B = \frac{\$1{,}000}{1.05} = \$952.4$$

This is a holding period return of:

$$\text{HPR} = \frac{952.4}{890} - 1 = 7\%$$

It appears as though the return from the longer term bond is higher than the return from the shorter term bond. This effect is known as riding the yield curve, since it is due to the fact that as time passes we are discounting the bond at lower yields.

However, that assumed that the term structure was unchanged. If the expectations hypothesis holds, then we should expect the differences between the 5% one-year yield and the 6% two-year yield to be due to expected changes in the one year rate. So we should expect the one-year rate one-year from now will not still be 5%.

We can determine the expected rate one year from now (assuming the expectations hypothesis holds) from the relationship:

$$1.06^2 = (1.05)[1 + E(r)]$$

and solve for $E(r) = 7\%$.

So we should really expect to pay \$890 today for the two-year bond and for it to be worth:

$$\frac{\$1{,}000}{1.07} = 934.5$$

at the end of the year. This results in a holding period return of:

$$\text{HPR} = \frac{934.5}{890} - 1 = 5\%$$

This is the same holding period return from investing in a one-year bond.

Question 8. List the two broad classes of passive bond fund management.

Solution. Indexing and Immunization

Question 9. List two approaches to eliminating interest rate risk in a bond portfolio and give an example of what type of institution would use each.

Solution. Two approaches are:

1. Protect total portfolio or net firm market value — used by banks to protect their net worth;

2. Protect expected future value with reinvestment — used by pension funds that have future payment obligations.

Question 10. List three reasons why the basic immunization approach described is unrealistic.

Solution. Three reasons are:

1. It assumes that interest rates only change by small amounts,
2. it assumes parallel shifts in the term structure and
3. it ignores inflation.

Question 11. List 2 types of bond swaps that are used when you think you have identified mispriced bonds and 2 that are used when you have a view about interest rate movements.

Solution. To take advantage of bonds that are mispriced, you can use a Substitution Swap or an Intermarket Spread Swap. To take advantage of specific views you may have about the direction of interest rate changes you can use a Rate Anticipation Swap or a Yield Pickup Swap.

Market Value

For simplicity, this paper separates the value of a P&C company into two components — Portfolio Equity and Franchise Equity.

Portfolio Equity

Portfolio Equity represents the net value of the currently booked assets and liabilities arising out of business already written. There are various ways to value these amounts and Noris presents three alternatives:

i. Book Value (Statutory Surplus) — Bonds are discounted according to their yields at the time of purchase, stocks are valued at market value and liabilities are not discounted at all.

ii. Current Value Surplus — Values all assets at current market value and the liabilities on an undiscounted basis.

iii. Market Value Surplus — This values assets at their current market value and also discounts the liabilities to a present value based on any number of possible interest rates, including the historic portfolio yield, the current yield of its assets, or some conservative figure. Noris uses the yield on municipal bonds — a proxy for the current after-tax yield.

Franchise Equity

Franchise Equity represents the value of business not yet booked. Because this is so difficult to value, this element of the firm value is ignored in this paper and is not part of the ALM strategy.

Managing Market Value Surplus

When insurers' surplus is stated on a market value basis, substantial variability in surplus becomes evident and suggests that it might be appropriate for insurers to manage this volatility. One mechanism for doing this is to manage the interest rate sensitivity of the market value surplus (MVS) by managing its (Macaulay) duration.

If we define the symbols D and MV to denote the duration and market values of the surplus, assets or liabilities, then:

$$D_{\mathrm{MVS}} = \frac{D_{\mathrm{MVA}}\mathrm{MVA} - D_{\mathrm{MVL}}\mathrm{MVL}}{\mathrm{MVS}}$$

Duration Gap of Surplus is then defined as the amount by which the duration of the MVS exceeds zero (zero indicating no sensitivity to interest rate changes). With a positive duration gap, MVS would fall as rates rise; a negative duration gap would indicate that MVS rises as rates rise.

Calculation Details

To calculate the duration gap, note that Noris uses Macaulay duration for the bonds. For stocks, he uses the following:

$$\text{Stock Duration} = \frac{1}{\text{Dividend Yield}}$$

Note that this stock duration measure assumes that the dividends themselves are not affected by changes in rates and ignores a number of other sources of common stock price variability other than changes in rates.

To see the details behind this assumption for the sensitivity of the stock price to changes in interest rates, note that under the simplistic no-growth dividend discount model, the value of the firm equity is given by:

$$V = \frac{Div}{k}$$

where Div is the (constant) dividend and k is the cost of capital or discount rate.

Taking the derivative with respect to k and assuming that the Div is not affected by changes in k, then:

$$\frac{dV}{dk} = \frac{d[Div(k^{-1})]}{dk} = -Div(k^{-2})$$

If we express this as a percentage of the value, $V = Div/k$, we get:

$$\text{Modified Duration} = \frac{-Div(k^{-2})}{V} = -Div(k^{-2})(k/Div) = -1/k$$

Then, if we ignore the negative sign as we did before when defining the duration for bonds, duration equals:

$$\text{Stock Duration} = \frac{1}{k} = \frac{V}{Div} = \frac{1}{\text{Dividend Yield}}$$

Note that this definition of duration is the *modified* duration, not the Macaulay duration, since it is merely the derivative with respect to the discount rate.

Alternative Derivation

Recall from calculus that we can write the *percentage* change in a function as the differences in the natural logs, and so:

$$\text{Modified Duration} = \frac{d[\ln V]}{dk} = \frac{d[\ln Div - \ln k]}{dk} = -\frac{1}{k}$$

Again ignoring the negative sign, this can be written as:

$$\text{Stock Duration} = \frac{1}{\text{Dividend Yield}}$$

Duration Gap Targets

Three different targets to manage towards are discussed:

i. Duration Gap of Surplus — One target for the duration gap of surplus (DG_S) is to set it to zero, thus immunizing the surplus from the effects of changes in interest rates. This might be desirable in some contexts but it would result in large fluctuations in earnings and would be unduly restrictive.

ii. Duration Gap of Total Return — Recall from BKM Chapter 16 that the Macaulay duration indicates the holding period over which the rate of return is immunized. Therefore, insurers might want to set their DG_S equal to a specific holding period (e.g. one year) and thereby achieve a target return on surplus over that period. If H is the desired holding period, then DG_{TRS} can be used to indicate the duration gap of the total return on surplus and then $DG_{\text{TRS}} = DG_S - H$.

iii. Duration Gap of Leverage — Others might be interested in the degree of leverage, which can be denoted as either MVS/MVA or its reciprocal. It should be clear that if both MVS and MVA change by the same percentage amount, then the leverage ratio will remain unchanged and will therefore be immunized against changes in rates. To do this, we simply set the duration of assets equal to the duration of surplus, since the (modified) duration tells us the percentage change in the surplus or assets. Noris then defines the duration gap of leverage as $DG_{EL} = D_{\text{MVS}} - D_{\text{MVA}}$.

Other Issues

The paper then discusses other issues facing P&C companies *at the time the paper was written.* He makes some points about the tax implications of discounting reserves (at the time a proposal but by now a fact of life) and about the dangers of cash flow underwriting. But he also touches on a topic that Feldblum discusses at length — the inflation sensitivity of losses.

Unexpected Loss Development

One source of unexpected loss development is unexpected changes in the inflation rate, since P&C liabilities are inflation dependent. But because fixed income assets have cash flows that are (usually) stated in nominal terms, duration strategies will not fully immunize against unexpected changes in inflation. What is needed is a way to ensure that investment returns keep pace with inflation.

Noris mentions three alternative approaches:

i. One way to do this is to invest more heavily in common stocks or real estate, two asset classes that arguably contain an exposure to inflation. Noris dismisses this because these asset classes introduce too much non-interest rate volatility.

ii. Another approach is to have assets that roll over frequently, but this will require more short-term assets and will likely require a deviation from duration matching.

iii. Another approach is to simply overstate the liabilities in the form of contingency reserves.

Practice Questions

Question 1. Suppose an insurer was following an asset-liability management strategy in which the Macaulay duration of its assets was equal to the Macaulay duration of its liabilities, both of which equaled 5. What is the Macaulay duration of their surplus?

Solution. This is sort of a trick question, because if you don't think about it you might quickly conclude that the surplus duration would be zero if you set the asset and liability durations equal to each other. But that assumes we were talking about the *dollar* duration. Using the normal units of measurement for Macaulay duration, we have the basic relationship from the reading:

$$D_{\mathrm{MVS}} = \frac{D_{\mathrm{MVA}}\mathrm{MVA} - D_{\mathrm{MVL}}\mathrm{MVL}}{\mathrm{MVS}}$$

If we note that MVS = MVA − MVL, then we can rearrange these terms slightly to write this as:

$$D_{\mathrm{MVS}} = \frac{\mathrm{MVA}}{\mathrm{MVS}}\left(D_{\mathrm{MVA}} - D_{\mathrm{MVL}}\right) + D_{\mathrm{MVL}}$$

From this, it should be obvious that when $D_{\mathrm{MVA}} = D_{\mathrm{MVL}}$ then the surplus duration will equal the liability duration, which in this case is equal to 5.

Note that had the asset and liability market values been equal, then zero surplus duration would have been achieved by simply setting the asset and liability Macaulay durations equal to each other. But this situation is unrealistic since it involves a company with no surplus value. Why hedge that to be immune to rate changes?

So when people talk about equating the asset and liability durations to hedge the change in surplus value, they are referring to the *dollar* durations (the Macaulay duration times the market value of the asset or liability).

Question 2. The current yield curve is flat with interest rates of 5% for all maturities. An insurance company has a liability payment of \$890.75 that will be paid in exactly six years, giving it a market value of $MVL = 664.69$. Their assets consist of a 5% coupon bond, with annual coupon payments, a \$1,000 face value and 6 years until maturity, giving the asset a market value of $MVA = 1,000$. The Macaulay duration of the liability is $D_{MVL} = 6.00$ and the Macaulay duration of the bond is $D_{MVA} = 5.3295$. If interest rates remain at 5% and coupon payments are reinvested, the value of the surplus (on an economic or market value basis) in four years will be 407.57. What would the surplus be in four years if instead interest rates were to suddenly rise to 5.05% and remain constant for four years?

Solution. Note that I tried to make this problem less distracting and tedious by giving you a lot of the values you need.

You were told the Macaulay durations for the asset and the liability, from which we can easily calculate the current surplus (on an economic or market value basis) as:

$$\begin{aligned}
\text{MVS} &= \sum_{t=1}^{6} \frac{50}{1.05^t} + \frac{1,000}{1.05^6} - \frac{890.75}{1.05^6} \\
&= 1,000 - 664.69 \\
&= 335.31
\end{aligned}$$

Knowing the various market values, we have the following for the surplus duration:

$$\begin{aligned}
D_{\text{MVS}} &= \frac{D_{\text{MVA}}\text{MVA} - D_{\text{MVL}}\text{MVL}}{\text{MVS}} \\
&= \frac{5.3295(1,000) - 6.00(664.69)}{335.31} \\
&= 4.00
\end{aligned}$$

We could also have calculated the surplus duration directly from the *net* cash flows:

TABLE 1. Surplus Duration

Time	$(1+r)^{-t}$	Cash Flow	PV(CF)	PV(CF)*t
1	0.9524	50.00	47.62	47.62
2	0.9070	50.00	45.35	90.70
3	0.8638	50.00	43.19	129.58
4	0.8227	50.00	41.14	164.54
5	0.7835	50.00	39.18	195.88
6	0.7462	159.25	118.83	713.01
			335.31	1,341.33
			Macaulay Duration	4.00

The question asked what the surplus would be in 4 years if rates increased a little bit, from 5% to 5.05%. Since the surplus duration is 4.00, we know that the holding period return on the surplus in 4 years is unchanged. So while the asset value and the liability value (at $t = 4$) both decline, they don't decline by the same amount, however the extra investment income from reinvesting the coupons at the higher rate offset this so that the surplus is the same as it would have been had rates stayed the same.

Another way to say the same thing is that the duration gap of the total return when the holding period is $H = 4$ is equal to zero. So we know that the surplus in four years (all other things held constant) is equal to \$407.57, the same as if rates were unchanged.

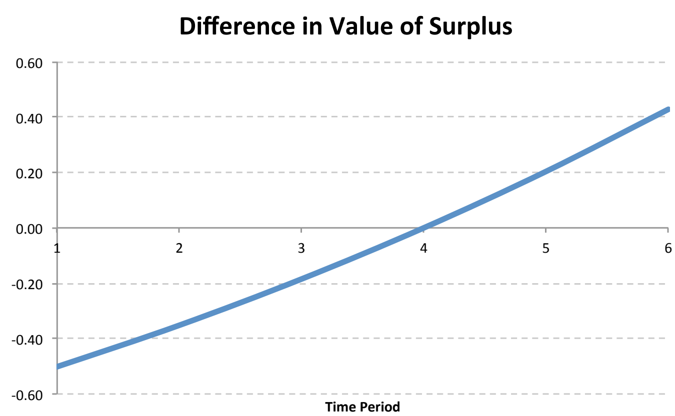
Here's a bit more of the calculations to help you see this key result. In the table, I show the surplus value at $t = 1, 2, \ldots$ after taking into account the reinvestment of the coupons. When rates stay the same at 5%, the surplus just grows at 5%. But when rates rise, at different holding periods the reinvestment of the coupons either increases by more or less than the change in the market values of the bond and the liability. Only at $t = 4$ is the result the same regardless of the interest rate.

TABLE 2. Market Value of Surpus at Different Horizons

	Rate Unchanged at 5.00%		Rate Changed to 5.05%	
Time Period	Value	Annual Return	Value	Annual Return
1	352.07	5.00%	351.57	4.85%
2	369.68	5.00%	369.33	4.95%
3	388.16	5.00%	387.98	4.98%
4	**407.57**	5.00%	**407.57**	5.00%
5	427.95	5.00%	428.15	5.01%
6	449.35	5.00%	449.77	5.02%

We can graph the differences in the surplus at different holding periods for the case where rates stay the same and the case when rates change:

FIGURE 1. Effect on Surplus of Change in Rates



Feldblum: Asset Liability Matching for P/C Insurers

Almost all of this paper is a review of material in BKM, so my comments below will focus primarily on the new material. Note that this discussion is largely about the strategy of Asset-Liability Matching and many of the observations are not necessarily generally applicable to Asset-Liability Management as more broadly discussed in BKM.

Difference Between Life and P&C Liabilities

Feldblum notes that Asset-Liability Matching strategies gained widespread popularity among life insurance companies to manage exposure to changes in interest rates, but that P&C operations and Life operations differ in ways that could impact the optimal asset-liability *management* strategy:

- Impact of Inflation Sensitivity on Effective Duration — Life liability cash flows, like traditional bond cash flows, are stated in nominal terms while P&C liability cash flows are inflation sensitive — as inflation rises the cash flows generally rise as well. As a result, Macaulay duration may not accurately reflect the interest rate sensitivity of P&C liability cash flows, since interest rate changes and changes in inflation are correlated. Note that this concept was discussed in BKM with respect to mortgage backed security cash flows and the discussion of effective duration.

- Impact of Duration Matching on Yield — Because P&C liability cash flows are generally of shorter duration than life cash flows, P&C insurers following an asset-liability matching strategy will have to invest in relatively short duration assets. With an upward sloping yield curve, this will mean more attention needs to be paid to the trade-off between the extra yield offered by longer duration bonds and the extra protection against interest rate sensitivity of surplus.

- Similarities Between Equities and P&C Liabilities — As Noris showed, the traditional Dividend Discount Model suggests that equity durations are quite long (20-30 years). As we'll discuss in a moment though, this is not really the case because it ignores the interest sensitivity of the equity cash flows. Equity cash flows and P&C liabilities share two similarities — both are inflation sensitive and both are subject to considerable risk other than interest rate risk (contagion and legal risk for the liabilities, systematic market risk for equities).

- Absence of Disintermediation Risk — Unlike life insurers, P&C insurers do not face disintermediation risk. This is the risk that customers of life insurers and other financial institutions will want to move their funds into more attractive investment products as rates rise. This is problematic because it requires them to sell assets at a time when their market values have declined.

- Mark-to-Market Risk of Long Term Bonds — Because life insurance liability cash flows occur in the distant future relative to P&C liability cash flows, life insurers can get away with focusing on GAAP earnings, which for long term bonds carried at amortized cost can be high and steady. But P&C insurers must be concerned with changes in market values because bond portfolios containing long-term bonds may need to be sold as the liabilities come due. This makes long term bonds riskier for P&C companies than for life companies.

Two Ways to Match Assets & Liabilities

Feldblum reviews the two methods of matching assets and liabilities (covered in BKM), which are Cash Flow Matching and Duration Matching.

- Cash Flow Matching — The three drawbacks to this simple strategy are repeated:

  i. it is cumbersome

  ii. it is inefficient

  iii. it is costly

- Duration Matching — The concept of Macaulay duration and the insensitivity of a bond's accumulated value (including reinvestment of coupons) to interest rate changes at the point in time equal to the Macaulay duration (the concept BKM referred to as Holding Period Immunization) are reviewed.

Loss Reserve Durations

Feldblum calculates the duration of P&C loss reserves using *reserve* payout patterns derived from Schedule P data. He assumes mid-year payments for each accident year, assumes payments after year 10 are paid evenly for five years, and uses a discount rate equal to the *current* yields on high rated corporate bonds. He finds that the reserve durations are approximately 3.2 years and notes that this is comparable to what others have calculated.

Note a few details:

- Payment Patterns — The payment patterns used are for the loss reserve payments, not the incurred loss payments. By ignoring the payments made in the year incurred, the reserve payment patterns are generally much slower than the incurred payment patterns.

- Discount Rate — The appropriate discount rate for calculating the liability duration is the current yield on investable assets with a similar maturity and risk profile as the liabilities. Since we are going to set the asset portfolio mix so as to match the liability durations, this is somewhat circular and requires an iterative approach. But Feldblum doesn't recommend "risk-free" rates and instead prefers Aaa-rated bond yields (or LIBOR in a more recent Feldblum paper).

- Asset Yield Used in Paper — The yield we want to use is the yield on new investments. Note though that to determine the 1983 liability duration, Feldblum used the yields in 1983, not the new money yields at the time of his paper. This is simply because he was trying to calculate what the durations would have been at the end of 1983, using the Schedule P information as if the 1983 payment patterns could have been known at the time.

Impact of Inflation

For liabilities that are inflation sensitive through the date of settlement, which probably applies to most P&C liabilities other than Workers Comp or Auto BI wage loss and medical bills, the increased nominal payments will largely offset the decrease in value caused by the higher discount rate (assuming that interest rate changes equal the inflation rate changes). This will cause the liabilities to have a duration of approximately zero — essentially the same as short term securities such as commercial paper or T-Bills.

So duration matching in this case will require a very short term asset portfolio or investments in inflation sensitive assets such as common stocks or real estate (or, as discussed in our coverage of BKM, in a combination of longer term bonds and securities with high negative durations such as mortgage backed IO securities).

Equity Duration

Since duration matching for P&C companies will require asset portfolios with short durations and relatively low yields, Feldblum recommends investing in inflation sensitive assets as an alternative. Since real estate is very illiquid, requires substantial expertise and is limited by regulatory restrictions, common stock seems to be an appropriate choice. This raises the question of what the duration of equity is.

Traditional View of Equity Duration

As discussed in Noris, if we use a simple Dividend Discount Model for the value of equity and ignore the interest rate sensitivity of the dividends, then equity duration can be found as:

$$\text{Modified Equity Duration} = \frac{1}{k-g}$$

where k is the cost of capital and g is the growth rate in dividends. As you can see, for typical values of of k and g this would produce equity durations of 10-20 or higher.

This can be derived just as we derived the similar formula in Noris, but there we assumed that $g = 0$. Here, with a growth rate of g, the value of the firm equity is given by:

$$V = \frac{Div_1}{k-g}$$

where Div_1 is the dividend, k is the cost of capital or discount rate and g is the growth rate.

Taking the derivative with respect to k and assuming that Div_i and g are not affected by changes in k, then we can take the derivative with respect to k, divide by the value to express the derivative as a percentage of the initial value and ignore the negative sign for consistency with our standard definition of duration. The result is the formula shown above.

Interest Rate Sensitivity of Cash Flows

To take into account the effect on the dividends (i.e. to calculate the *effective* durations of the equity), Feldblum assumes that interest rates and inflation rates are correlated and then notes the following factors causing common stock prices to fall in the short run but then rise in the long run in response to an increase in rates:

- Value of Assets — If interest rates rise because of changes in inflation (and note carefully that rates can rise for different reasons as well), then the value of real assets should also rise and the inflation adjusted value of the assets should be insensitive to the changes in the rates in the long run.

- Demand Pull vs. Supply Push — Inflation can be caused by two sorts of phenomena. If it is *demand pull* then that means that excess demand has made people willing to pay more for now scarce product and this should coincide with more firm revenues. If it is *supply push*, firms are merely raising price to pass along their increased costs and this may lower demand. Furthermore, higher interest rates may result in increased savings and lower demand. So inflation has an ambiguous effect on demand. But it has an unambiguous effect on supply costs and so inflation may cause profits to either increase or decrease. If demand does not increase, then profits will fall and the value of the dividend, and hence the common stock, will fall.

- Market Demand for Stocks — As rates rise, investors may shift their holdings to bonds versus stocks and thus overall demand for stocks may fall, causing stock prices to fall.

The first effect is a long term effect but the latter two are short term. Hence, Feldblum concludes that as rates rise, stock prices will first decline and then rise. He then shows that stock prices are negatively correlated with inflation but are positively correlated with inflation lagged one period.

Other Considerations

Managing interest sensitivity necessarily involves trade-offs and is not the only consideration. Others include:

- Yields — Shortening the asset duration to match the liability duration, if done with a bond portfolio, will in most instances involve giving up yield because long term yields typically are higher than short term yields. Note though that there are a lot of possible explanations for this. If the yield curve is upward sloping because short term yields are expected to rise, then there may be little trade-off. If long term yields are higher because of liquidity premiums, then firms with greater liquidity needs cannot really afford to take these risks and it is not clear (to me at least) that there really is a trade-off (you aren't giving up something by not taking a risk that you shouldn't be taking anyway).

- Transaction Costs — These are lowest for long term bond portfolios, especially under a buy and hold strategy, and are considerably higher for common stocks — both the trading costs and the costs to analyze securities.

- Disintermediation — Feldblum notes that technically speaking, disintermediation doesn't really affect the value of an insurer — it merely forces them to recognize the loss in value. For example, if rates rise and the policyholder of a whole life policy with a policy loan provision wants to take the loan, then the insurer may need to liquidate its bond portfolio (to come up with the cash) at a loss. But the loss already occurred when the rates rose and had the portfolio been marked to market it would have had to be recognized.

- Cash Flow Risk — Feldblum argues that a P&C insurer is unlikely to become insolvent just because rates rise, in part because long term bonds are held at amortized cost on the statutory balance sheets and in part because they hold a single asset account and can use current premium to pay claims. Therefore, he argues that a P&C insurer can make yield a higher priority than duration matched assets and liabilities — offering two options, a short duration common stock portfolio with exposure to systematic market risk or a long term bond portfolio which exposure to mark to market gains and losses.

Practice Questions

Question 1. Feldblum says that both property and liability reserves have low durations, but for different reasons. What are the two reasons?

Solution. Liability reserves are inflation sensitive and thus have interest rate sensitivity comparable to low duration assets. Property reserves are not inflation sensitive but are paid out quickly, so they too have low durations.

Question 2. Suppose that the risk free rate is currently 5%, the equity risk premium is 8% and the expected dividend growth rate for the S&P 500 is 3%. What is the duration of the S&P 500 index under the assumptions that dividends grow at a constant rate in perpetuity, the CAPM is an appropriate model of expected equity returns and that dividends are not interest rate sensitive?

Solution. Under the standard DDM, and ignoring the interest rate sensitivity of the dividends themselves, then the duration of equity is equal to $1/(k - g)$. In this case, the beta for the S&P 500 can be assumed to equal 1.0 and $k = 5\% + 1(8\%) = 13\%$ and $g = 3\%$, so:

$$\text{Duration of S\&P 500} = \frac{1}{.13 - .03} = 10$$

Question 3. Contrast the impact of changes in inflation rates on the value of a retail store with the impact on the value of a regulated utility.

Solution. When inflation rates rise, retail stores are able to increase their prices accordingly and thus its nominal value will rise with inflation and its present value should be largely unchanged. However, a regulated firm cannot raise prices as easily and thus inflation will cause its costs to rise but revenues to remain unchanged, thus lowering its value.

Panning: Managing Interest Rate Risk

This paper begins by demonstrating the role of *franchise value* in measuring and managing the degree to which the economic value of an insurer is adversely exposed to changes in interest rates. It then goes on to suggest that a firm can manage the overall impact of interest rate changes through its pricing strategies and discusses the advantages this may have over traditional methods that rely on varying the asset investment strategies instead.

Franchise Value

When we refer to the *value* of a firm, it is common to refer to the *accounting value* of the surplus, which values the assets currently held and then subtracts the reported existing liabilities. As we will see in the Valuation section of these notes, this calculation will rarely be the same as the market value of the firm's equity. Aside from the failure to reflect the current economic value of the existing liabilities (reflecting discounting for the time value of money and the inclusion of a risk margin), the accounting value does not take into account the value of future business.

If our goal is to manage the interest rate risk to the whole firm, then it is important to include the franchise value. It is not sufficient to simply measure and manage the interest rate risk of the assets and liabilities recognized on the balance sheet. We should instead manage the interest rate risk that the total (market) value of the firm is exposed to.

Panning notes three important points about franchise value:

i. Franchise value is significant for many P&C insurers. For instance, in 2005 the average ratio of the market value to book value for the P&C industry was 1.54. The franchise value is at least a third of the total value of the firm and therefore is not easy to ignore.

ii. Franchise value is exposed to interest rate risk. Since the franchise value reflects the present value of the future cash flows from renewals (and "new" business), its value will vary as interest rates change.

iii. Franchise value is often unmeasured, unreported and consequently unmanaged by the firm.

Simple Model

Panning uses a very simple model to illuminate the key insights of the paper. Don't let this simplicity fool you though — the conclusions drawn are broadly applicable in more complex settings. His key assumptions of this simple model are:

- All business written in Jan 1
- All expenses paid immediately

- All claims are paid in full at the end of one year
- Expenses and expected losses are the same each year
- It maintains the same surplus each year. It does this by paying out all profits at the end of each year and raising new capital (with no frictional costs) to cover any net losses.
- It faces no risk of bankruptcy — surplus is sufficient to cover any deviation of losses from their expected value
- There are no taxes
- Term structure of interest rates is flat
- All calculations done on Jan 1

Notation

Panning uses the following notation:

- P = Premium
- E = Expenses in dollars; assumed to be the same each year with 100% client retention
- L = Expected claims and loss adjustment expenses, in dollars; assumed to be the same each year with 100% client retention
- y = risk-free interest rate
- S = Surplus, assumed to be constant
- k = Target return on surplus
- cr = Client retention rate, reflecting the percentage of clients who renew each year
- F = Franchise value reflecting the present value of net cash flows from future retentions
- C = Current economic value of assets and liabilities on the balance sheet

Policy Pricing

Note that Panning assumes the firm prices its policies such that it expects to earn a return on surplus of k. As a result, the sum of the premiums, losses, expenses and investment income on the surplus and net premium must equal k times the initial surplus:

$$P - L - E + (S + P - E)y = kS$$

Solving for P, we have the following formula for the premium that would be charged given these other assumptions:

$$P = \frac{S(k - y) + L}{1 + y} + E$$

Firm Value and Franchise Value

Noting the formula above for the premium, we can easily calculate the current economic value (immediately after writing the business on Jan 1) of the surplus and the net cash flows from the business. This is just the surplus, plus premium, less expenses, less the risk-free present value of the claims (for simplicity Panning ignores risk margins on the losses and simply discounts the expected loss at the risk-free rate):

$$\text{Current Economic Value} = \text{Surplus} + \text{Premium} - \text{Expenses} - \text{PV(Losses)}$$

$$C = S + P - E - \frac{L}{1 + y}$$

To calculate the franchise value, we need only recognize that with the simplifying assumptions used here the value of the future premiums, expenses and claims are simple perpetuity calculations, with all amounts decreasing by the ratio cr each period. So, for the first *future* period, the premium is $P(cr)$ and the present value of this is:

$$\begin{aligned} PV(P_1) &= \frac{P(cr)}{1 + y} \\ &= Pd \qquad \text{where } d = \frac{cr}{1 + y} \end{aligned}$$

Then, for the next future period the present value (today) is $P(d^2)$. The total value of this perpetuity converges to:

$$PV(\text{Future Premium}) = \frac{Pd}{1 - d}$$

The expenses are handled identically. The claims are also handled essentially the same way, but because they are paid at the end of the year, there is one extra year of discounting. The net effect is a simple formula for franchise value:

$$F = \left[P - E - \frac{L}{1 + y}\right]\left[\frac{d}{1 - d}\right]$$

Interest Rate Sensitivity of Franchise Value

Using the formula above, it is easy to determine how sensitive the franchise value is to changes in interest rates. However, there is one modification that is introduced first to ensure that the formulas fully account for this interest rate sensitivity.

In the formula above, the firm's target return on surplus was assumed to be a constant k. Panning introduces the concept of a *pricing strategy* that allows this target to vary based on the current interest rates. To be as flexible as possible, he sets k equal to a linear function of the risk-free rate:

$$k = a + by$$

and then allows a and b to be parameters that are set to determine the degree to which current interest rates alter the firm's target return and hence their pricing. These parameters together account for the firm's pricing strategy.

Introducing this adjustment makes the franchise value formula messier, as shown below:

$$F = \frac{(cr)S(a + (b-1)y)}{(1+y)(1+y-cr)}$$

But now when we determine the duration (by taking the derivative with respect to y) we are fully accounting for the effective duration of the franchise value, including the impact on the premium cash flows (the expense and liability cash flows are assumed to not be interest sensitive).

The steps involved are tedious, but the end result is the following estimate for the effective duration of franchise value:

$$D = \frac{a-b+1}{(1+y)(a+by-y)} + \frac{1}{1+y-cr}$$

Table 1 in the text displays a numerical example using specific values for the key assumptions. In this example, the assumptions are:

$$S = 50, L = 75, E = 25, cr = 90\%, a = 15\%, b = 0, y = 5\%$$

This produces franchise value duration of 17.62, as shown in Table 1. Notice in the table

TABLE 1. PV and Duration of Franchise Value

	Annual Value	PV	Duration	Dollar Duration
Premiums	101.19	607.14	7.85	4,768.71
Losses	-75.00	-428.57	7.62	-3,265.31
Expenses	-25.00	-150.00	6.67	-1,000.00
Total		28.57	17.62	503.40

that each of the components are calculated separately to highlight the fact that the premium duration is higher than the liability duration, even though premiums are collected one year earlier.

The reason for this is that the premium cash flows are interest sensitive, especially in the case here. Based on the pricing strategy parameters a and b used in the above example, the target return on capital is fixed. Recall the formula for the premium:

$$P = \frac{S(k - y) + L}{1 + y} + E$$

If k is fixed at 15%, then as interest rates rise the premium cash flows fall. This makes the interest sensitivity much greater than it would be for a fixed cash flow stream.

Managing the Interest Sensitivity of Franchise Value

Given the one year maturity of the firm's liabilities, their liability duration will be approximately one. If the firm invested in assets with a duration of one, then it would be natural to refer to this firm as being duration matched. Recall from the Norris reading that this does not mean that its surplus is immunized against changes in interest rates, because there are more assets than liabilities and so the *dollar durations* are not equal. But the overall surplus duration will be approximately one. The firm may find this acceptable or may choose to reduce it, but the magnitude does not appear to be too troublesome.

However, this ignores the fact that the franchise value may be significant and its duration, as well as its dollar duration, can be quite large (as in the specific example referred to above). In the example in the text, the current economic value is \$54.76 and has a duration of approximately 1.0. But the franchise value is \$28.57 and has a duration of 17.62. The overall duration of the firm's total economic value is then 6.70.

Notice that if we wanted to reduce this duration (i.e. reduce the impact of changes in interest rates on the firm's shareholders), the traditional options are to reduce the duration of the invested assets or to purchase derivative securities with high negative duration. But in this particular example, even reducing the duration of the assets to zero would only reduce the duration of the firm's economic value to 5.18.

Notice two critical points:

1. Limits of Using Investment Strategies to Manage Interest Rate Risk — The larger the firm's franchise value, the more problematic it will be to manage the interest rate risk for the total economic value using asset investment strategies.

2. Invisibility of Franchise Value — Because the franchise value is *invisible* to outside parties such as rating agencies and regulators (as well as investors), strategies to manage this risk may actually appear to increase the risk. The figures shown on the balance sheet will suggest a large duration mismatch and the complex, negative duration assets will likely appear risky.

As a result, it is useful to consider whether there are other ways to manage the interest rate risk associated with the firm's franchise value.

Using Pricing Strategy to Manage Duration of Total Economic Value

Panning's suggested solution to this dilemma is to rely on the firm's pricing strategy to manage the interest rate sensitivity of franchise value. In the examples shown earlier, the firm's target return on surplus was set to a constant of 15% by setting $a = 15\%$ and $b = 0$ in the formula for the target return on surplus $k = a + by$.

Suppose that we change these parameters to $a = 10\%$ and $b = 1$, so that at the assumed risk-free rate of $y = 5\%$ the target return is still $k = 15\%$. However, now when rates change and we allow our target return to change, the duration of franchise value drops from 17.62 to 7.62. This causes a decline in the duration of total economic value from 6.70 to 3.27.

As with all uses of duration to manage interest rate risk, it may be difficult to maintain a rigid relationship between the target return on surplus and the target duration of total economic value for anything other than very small changes in interest rates. However, this is a weakness common to all duration-based risk management strategies.

Panning notes a key virtue of this approach:

It avoids the potential rating agency and regulatory risk associated with strategies that focus on managing the duration of the firm's invested assets as a means of managing the risk to its franchise value and total economic value. This key advantage results from the fact that implementing a pricing strategy is nearly as invisible to these external audiences as the franchise value it is intended to protect.

Practice Questions

Question 1. Suppose an insurer is established with initial surplus of $S = 50$. They write policies on January 1 that have up-front expenses of $E = 25$ and expected claims that will be paid in full at the end of the year equal to $L = 75$. They have a target return on surplus of $k = a + by$ with parameters $a = 10\%$ and $b = 1$. Use Panning's premium formula to determine what they would charge if the current risk free rate is 5%.

Solution. Panning uses the following formula for the premium:

$$P = \frac{S(k - y) + L}{1 + y} + E$$

In this case, the pricing strategy parameters a and b suggest a target return on surplus of:

$$k = 10\% + 1(5\%) = 15\%$$

This leads to premium equal to:

$$P = \frac{50(.15 - .05) + 75}{1.05} + 25 = 101.19$$

Question 2. Immediately after writing the policies in the previous problem, what is the current economic value of the firm's surplus and the duration of surplus assuming that all of the assets are invested in risk-free securities with a duration of 1.0? For simplicity, ignore risk margins when determining the present value of the liabilities.

Solution. Immediately after writing the policies, the current economic value is:

$$
\begin{aligned}
C &= S + P - E - \frac{L}{1+y} \\
&= 50 + 101.19 - 25 - \frac{75}{1.05} \\
&= 54.76
\end{aligned}
$$

For the duration, note that the invested assets of the firm will be:

$$A = 50 + 101.19 - 25 = 126.19$$

Because the asset duration was assumed to be equal to 1.0, the dollar duration of the assets is then 126.19.

The present value of the liabilities is $75/1.05 = 71.43$ and the modified duration is:

$$D_L = \frac{1}{1+y} = .9523$$

resulting in a dollar duration for the liabilities of $(71.43)(.9523) = 68.027$. The dollar duration of surplus is then 58.16 and the duration of surplus is:

$$D_S = \frac{58.16}{54.76} = 1.062$$

Note that in the reading the duration of the liability is referred to as simply being approximately 1.0. Using 1.0 rather than 0.9523 would cause the numbers to differ slightly.

Question 3. Assume that the firm in the previous questions will write the same business in the future in perpetuity, but the volume will be reduced each year according to a retention rate of $cr = 95\%$. What is the franchise value for this firm? What is the total economic value of the firm?

Solution. The franchise value is the present value of the net premium, expense and claim cash flows in perpetuity. While you can calculate each component separately, you can also use the following substitution:

$$d = \frac{cr}{1+y} = \frac{.95}{1.05} = 0.90476$$

and calculate the total franchise value as:

$$
\begin{aligned}
F &= \left[P - E - \frac{L}{1+y}\right]\left[\frac{d}{1-d}\right] \\
&= [101.19 - 25 - 71.43]\left[\frac{0.90476}{1-0.90476}\right] \\
&= 45.24
\end{aligned}
$$

The total economic value is therefore $54.76 + 45.24 = 100$.

Note that you could also use Panning's other formula for the franchise value after substituting in for P and d in terms of y:

$$F = \frac{(cr)S(a + (b-1)y)}{(1+y)(1+y-cr)} = 45.24$$

Question 4. What is the duration of the firm's franchise value using Panning's formula?

Solution. Taking the derivative of the franchise value formula with respect to the risk free rate, y, is tedious, but when that is done the duration of franchise value turns out to be:

$$D = \frac{a-b+1}{(1+y)(a+by-y)} + \frac{1}{1+y-cr} = 10.952$$

Question 5. What is the duration of the firm's total economic value?

Solution. Given the current economic value, the franchise value and their durations, the dollar duration of the total economic value is:

$$\begin{aligned} \text{Dollar Duration} &= (1.062)(54.76) + (10.952)(45.24) \\ &= 553.63 \end{aligned}$$

Dividing by the total economic value, the duration is:

$$D = \frac{553.63}{100} = 5.5363$$

Note that in the reading the duration of the liability is referred to as simply being approximately 1.0. Using 1.0 rather than 0.9523 would cause the duration of the surplus to differ slightly from the 1.062 value calculated earlier.

Question 6. What would happen to the duration of total economic value if the duration of the invested assets were set to zero?

Solution. In this case, the duration of total economic value would become 4.27.

Question 7. What if instead of altering the asset duration the firm changed its pricing strategy such that $a = 5.28\%$ and $b = 1.945$. What would be the duration of the total economic value?

Solution. Notice that these parameters would still produce a target return on surplus of $k = 15\%$ and the same franchise value. However, the duration of the franchise value would decline to

$$D = \frac{a-b+1}{(1+y)(a+by-y)} + \frac{1}{1+y-cr} = 1.507$$

Using the result from Question 2 for the surplus duration of 1.062, the duration of the total economic value would be equal to:

$$\begin{aligned}\text{Duration} &= \frac{(1.062)(54.76) + (1.507)(45.24)}{100} \\ &= 1.264\end{aligned}$$

This is significantly lower than the duration that can be achieved by simply reducing the asset duration to zero.

Question 8. Panning suggests that rather than altering the composition and duration of the invested asset portfolio firms could use their pricing strategy, as defined by how sensitive their target return on surplus is to changes in the risk free rates, to effectively manage the interest sensitivity of the total economic value of the firm, including the franchise value. What does he say is key virtue of this approach with respect to rating agencies and regulators?

Solution. The concern is that the strategies that would have to be used to manage the duration of the total economic value through changes in the asset duration would likely lead to apparent duration mismatch and to the use of complex derivative securities that would possibly even appear to increase risk. This occurs because the franchise value is invisible to the rating agencies and regulators and therefore its effect is ignored.

A key advantage of using pricing strategy instead results from the fact that implementing a pricing strategy is nearly as invisible to these external audiences as the franchise value it is intended to protect.

Selected Old Exam Questions for Part 3

The following questions relevant for this section appeared on the Old CAS Exam 8 from 2000 to 2010 and on the CAS Exam 9 since 2011.

BKM 15	BKM 16	Hull 4	Hull 7	Noris	Feldblum ALM	Panning
2000 Q16	2000 Q14	2001 Q27	2003 Q34	2000 Q40	2000 Q39	2007 Q29
2001 Q16	2000 Q2	2002 Q24	2003 Q35	2001 Q42B	2001 Q42A	2008 Q32
2001 Q19	2000 Q43	2006 Q11	2003 Q9	2001 Q42C	2001 Q42D	2010 Q24
2003 Q25	2001 Q15	2006 Q12	2005 Q25	2002 Q34	2002 Q37	2012 Q5
2003 Q26	2001 Q18	2007 Q10	2006 Q20	2005 Q33	2004 Q33	2013 Q9
2003 Q27	2001 Q23	2008 Q12	2007 Q18	2007 Q27	2007 Q28	2014 Q10
2004 Q15	2002 Q17	2008 Q30	2007 Q19	2009 Q25	2009 Q26	
2004 Q16	2002 Q21	2009 Q6	2008 Q20	2011 Q6	2010 Q21	
2004 Q17	2003 Q6	2010 Q7	2008 Q21	2013 Q8	2012 Q4	
2005 Q18	2003 Q8	2012 Q2	2009 Q18		2014 Q11	
2009 Q10	2004 Q32	2014 Q7	2010 Q11		2015 Q10	
2009 Q9	2005 Q34	2015 Q8	2012 Q6			
2014 Q8	2006 Q31		2013 Q7			
	2006 Q32		2014 Q9			
	2008 Q31		2015 Q9			
	2009 Q27					
	2010 Q22					
	2011 Q5					
	2011 Q7					
	2011 Q8					
	2013 Q6					
	2015 Q11					

For some of these questions I have provided the text of the question and an explanatory solution. These were selected either because they are representative of the questions you are likely to be asked on future exams or because they contain an element that is particularly worthwhile to point out. For the other questions, the CAS solutions should be sufficient to confirm whether your answer is correct.

Important Note: The solutions shown here are intentionally detailed. They contain thorough explanations of the concepts and formulas used to reinforce the main points from the readings and provide an additional teaching opportunity. **Actual exam responses should be much more concise than what is shown here, along the lines of what you will see in the solutions that the CAS releases.**

2003 Exam Question 25

The following is a list of prices for zero coupon bonds with par value of \$1,000 and varying maturities.

Maturity (Years)	Price
1	943.40
2	898.47
3	847.62
4	792.16

a. Calculate the third year forward rate.

The term "third year forward rate" is potentially misleading, but it refers to the forward rate *in the third year* — which is the rate from $T = 2$ to $T = 3$.

We can find this by first determining the yield for a two year zero coupon bond and a three year zero coupon bond.

Using the information given, and assuming annual compounding, we can solve for the two-year and three-year spot rates as follows:

$$898.47 = 1{,}000(1 + y_2)^{-2} \Rightarrow y_2 = 5.499\%$$

$$847.62 = 1{,}000(1 + y_3)^{-3} \Rightarrow y_3 = 5.6654\%$$

Using these two yields the forward rate is such that investing for three years provides the same total return as investing for two years at y_2 and then investing at the forward rate f_3.

Therefore,

$$(1 + y_3)^3 = (1 + y_2)^2(1 + f_3)$$

and then solve for $f_3 = 5.999\%$.

b. Calculate the yield to maturity for the three-year bond.

This was calculated above as 5.6654%.

2003 Exam Question 6

A newly issued bond has maturity of 3 years and pays a 7% coupon rate (with annual coupon payments). For the questions below, assume that the bond was purchased at par ($100).

a. Calculate the modified duration and convexity of the bond.

Since the bond was issued at par, we know that the yield is 7% (as indicated in Part B).

Modified duration is calculated as follows by first calculating the weighted average time to payment (Macaulay duration) and then calculating Modified Duration as Macaulay Duration$/(1 + y)$.

Period	T	Cash Flow	PV(CF)	T*PV(CF)
1	1	7	6.54	6.54
2	2	7	6.11	12.23
3	3	107	87.34	262.03
			100.00	280.80
			Macaulay Duration	**2.808**
			Modified Duration	**2.624**

Convexity is calculated similarly using the formula from the footnote in the text,

$$\text{Convextity} = \frac{1}{P(1+y)^2} \sum \left[\frac{CF_t}{(1+y)^t} t(t+1) \right]$$

This is shown using most of the same values as above:

Period	T	Cash Flow	PV(CF)	(T+1)*T*PV(CF)
1	1	7	6.54	13.08
2	2	7	6.11	36.68
3	3	107	87.34	1,048.13
			100.00	1,097.90
			Convexity	**9.589**

Note that you could have also used the shortcut formulas for the duration and convexity approximations shown in my notes to save a little bit of time, though since there were only three cash flows the above answer was no doubt safer. If you did use the shortcut, it might be wise to alert the graders that these shortcuts are valid approximations discussed in BKM (under the heading Effective Duration).

b. Given that the bond's yield to maturity immediately increases from 7% to 8% with the maturity still 3 years, what is the percentage change in the price of the bond?

The bond's initial price was 100. It is easy to calculate the price at the new 8% yield to be 97.42 and so the percentage price change is −2.58%.

c. Using the duration rule, including the convexity adjustment, what is the percentage change in the bond price?

The "duration rule, including the convexity adjustment" says that the percentage change in price is as follows:

$$\begin{aligned}\frac{\Delta P}{P} &= -\text{Modified Duration}\,(\Delta y) + \frac{1}{2}\text{Convexity}\,(\Delta y)^2\\ &= -2.624(1\%) + .5(9.589)(1\%)^2\\ &= -2.576\%\end{aligned}$$

2002 Exam Question 17

You are the holder of the following bond:

- Term to maturity is 20 years
- 8% coupon, paid semi-annually
- Yield to maturity is 10%
- Par value is $1,000

You want to test this bond's price sensitivity to a 200 basis point decrease in interest rates.

a. Using the duration approximation, calculate the estimated price following the 200 basis point interest rate decline and determine the error in this approximation.

The first step is to calculate the duration of this bond. However, on the exam you cannot realistically calculate the modified duration directly using the formula in the BKM textbook when you have 20 years of semi-annual cash flows. Instead, you can use the formula that BKM call "effective duration" that they use to calculate the duration of interest-sensitive cash flows:

$$\text{Modified Duration} = \frac{P_- - P_+}{2P\Delta y}$$

This formula was presented in my notes as a valid *approximation* to modified duration. It used to appear directly on the syllabus back in 2002 when this question was asked on the exam.

We start with the value of the bond. Using your calculators, it is easy to find the bond price as $P = 828.4$.

Then if $\Delta y = 10$ bps, $P_- = 835.86$ and $P_+ = 821.05$.

We can approximate modified duration using the formula above:

$$\begin{aligned}\text{Modified Duration} &= \frac{P_- - P_+}{2P\Delta y} \\ &= \frac{835.86 - 821.05}{2(828.4)(.001)} \\ &= 8.94\end{aligned}$$

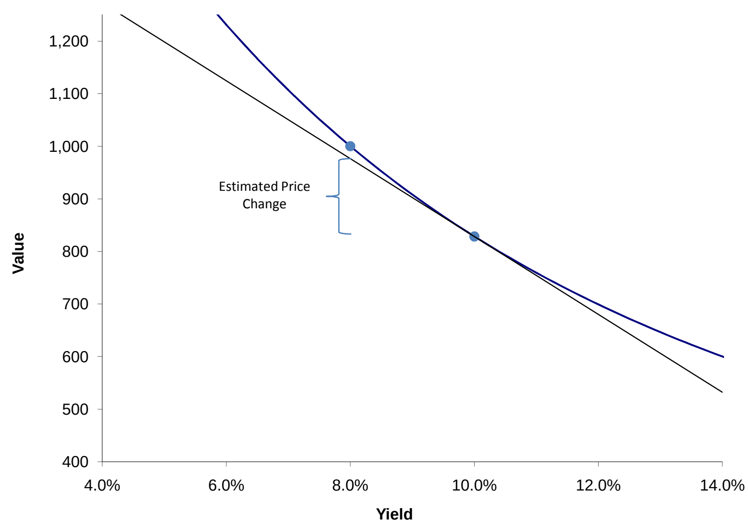
Then, to approximate the percentage change in price for a 200 basis point decline in the yield, we use the equation:

$$\begin{aligned}\text{Approx \% change in price} &= -D^* \Delta y \\ &= -8.94(-2\%) \\ &= 17.87\%.\end{aligned}$$

The new price would then be approximately $828.4(1.1787\%) = 976.43$.

The actual new price would be found using your calculator as \$1,000, which was obvious because the yield fell to be equal to the coupon rate. The error in the approximation is therefore \$23.57.

b. Graph the price/yield relationship for this bond and illustrate the duration approximation error calculated in Part (a) above.



c. Use the convexity adjustment to calculate a more accurate estimate of the price following a 200 basis point interest rate drop and determine the error in this new approximation.

From my notes, convexity can be approximated using the formula:

$$\text{Convexity} = \frac{P_- + P_+ - 2P}{P(\Delta y)^2} = 132.78$$

Then the convexity adjustment to the approximate change in price would be to add the quantity:

$$\frac{1}{2}\text{Convexity}\,(\Delta y)^2 = .5(132.78)(.02^2) = 2.66\%$$

to the approximate change calculated earlier. This gives a total change in the bond price as 17.87% + 2.66% = 20.525%.

The estimated price would then be 828.4(1.20525) = 998.44. This is much closer to the true value, with an error of only $1.56.

2007 Exam Question 18

You are given the following information about interest rates for Company A and Company B:

	Company A	Company B
US Dollars (floating rate)	LIBOR + 1.5%	LIBOR + 2.5%
Canadian Dollars (fixed rate)	7.50%	10.00%

Company A wants to borrow U.S. dollars at a floating rate of interest while Company B wants to borrow Canadian dollars at a fixed rate of interest.

To arrange a swap that is equally attractive to both Company A and B, a financial institution requires a spread of 25 basis points.

Determine the terms of the swap that could be entered into and the rates of interest Company A and Company B will each pay, after consideration of the swap.

The first thing to notice is the Company A can always borrow at lower rates than Company B, but the advantage in U.S. dollars is only 1.0% while the advantage in Canadian dollars is 2.5%. This gives Company A a *comparative advantage* of 1.5% if they borrow in Canadian dollars. If they were to enter into a swap with Company B, they could find a way to split this comparative advantage.

Of course, the financial institution wants to earn 25 basis points (.25%), so that leaves 1.25% for Company A and Company B to split. They will of course negotiate how this is split, but here they wanted the swap to be equally attractive, suggesting they split it evenly .625% each.

Company A can either borrow at 7.5% in Canadian dollars at a fixed rate or equivalently borrow at LIBOR + 1.5%. Since Company A has a comparative advantage in Canadian dollars he will borrow on that basis and then enter into a swap to convert his effective rate into a U.S. dollar floating rate. He'll pay 7.5% on the borrowing but has to wind up saving .625% from the swap. The swap will therefore have to consist of him receiving 7.5% (to offset his borrowing rate) and only paying LIBOR + .875% (which is 0.625% lower than if he had borrowed directly). His net payments will therefore be LIBOR + 0.875%.

Note that the swap above has Company A receiving 7.5% and paying LIBOR + 0.875%. They will pay and receive those amounts from the financial institution that takes .25%, so Company B's swap will be the same terms net of the .25% fee. That means, for Company B the swap will involve paying 7.5% fixed and receiving LIBOR + 0.625%.

To see what their net payments are, they have to borrow at LIBOR + 2.5%, so they wind up paying on a net basis a fixed rate of 7.5% + LIBOR + 2.5% − [LIBOR + 0.625%] = 9.375%.

Both parties wind up effectively borrowing in their desired currencies (Company A at U.S. dollar floating rate and Company B at Canadian dollar fixed rate) but save 0.625% compared to what they would have paid on those terms without the swap.

2007 Exam Question 19

A financial institution pays 6-month LIBOR and receives 7% per year with semi-annual compounding on a swap with notional principal of \$50 million. The remaining payment dates are in 4, 10, and 16 months.

The LIBOR rates with continuous compounding for 4, 10 and 16 month maturities are 8%, 8.5% and 9% respectively.

The 6-month LIBOR rate at the last payment date was 8.8% with semi-annual compounding.

Calculate the value of the swap to the financial institution.

Here the financial institution is paying a floating rate and receiving a fixed rate. The value of the swap to them is then the value of a long position in a fixed rate bond paying 7% semi-annually and a short position in a floating rate bond.

The fixed rate bond is trivial to value, since the series of cash flows is defined and the LIBOR rates are specified. The cash flows consist of 3.5% times \$50 million on each of the three dates and \$50 million at maturity.

The floating rate bond will be worth par on the next reset date, but the rate was set at 8.8% on the last reset date, so we need to value the known 4-month cash flow using the current LIBOR rates for that maturity. The 4-month cash flow is effectively 4.4% times \$50 million plus par value of \$50 million.

These calculations are shown below:

			Fixed Rate		Floating Rate	
Maturity	LIBOR	Present Value	Cash Flows	Value	Cash Flows	Value
4	8.00%	0.9737	1.75	1.70	52.20	50.83
10	8.50%	0.9316	1.75	1.63		0.00
16	9.00%	0.8869	51.75	45.90		0.00
				49.23		50.83
					Net Value	**-1.59**

Note that the more involved solution is to calculate each of the forward rates, convert those to semi-annual rates, assume those rates are realized and then discount the net cash flows at the continuously compounded LIBOR rates given.

2001 Exam Question 27

a. Briefly describe a forward rate agreement (FRA).

An FRA is simply an agreement between two parties for one to pay to the other a specific rate of interest, R_k, during some future time period on a specified notional amount L. The rate applies to a future time period, starting at T_1 and ending at T_2, and is expressed as an annual rate, even though the actual payment is adjusted to reflect the length of the time period.

b. You are given the following information:

- The two-year zero-coupon interest rate is 7.5%, compounded continuously.
- The three-year zero-coupon interest rate is 8%, compounded continuously.

You entered into a forward-rate agreement some time ago where you agreed to receive 10% with annual compounding on a principal of $2 million between the end of what is now year two and the end of what is now year three.

What is the value of the FRA now?

Before determining the value of the FRA, note that you could always lock in the forward rate, R_F, at no cost by borrowing the present value of L until T_1 and investing the same amount until T_2. In this case, $T_1 = 2$ years and $T_2 = 3$ years.

Assume (to keep things as simple as possible) that $L = 1$. In this case, you would borrow $1e^{-.075(2)}$ for two years and repay $1 at $T_1 = 2$ years.

At the same time, you would invest that same amount for 3 years and at $T_2 = 3$ years you would get back the amount you borrowed with interest, which in this case equals $[1e^{-.075(2)}]e^{.08(3)}$.

The net cash flows are therefore 0 now, $1 at T_1 and $[1e^{-.075(2)}]e^{.08(3)}$ at T_2. The rate of return that you earn from T_1 to T_2 , the forward rate, is therefore:

$$e^{R_F(1)} = \frac{[1e^{-.075(2)}]e^{.08(3)}}{1} = e^{.08(3)-.075(2)} \Rightarrow R_F = \frac{.08(3) - .075(2)}{3-2} = .09$$

To determine the value of an FRA, we simply note that the FRA locked in a rate of 10% on an annually compounded basis and we could always lock in 9% on a continuously compounded basis at no cost. So the FRA must be worth the difference in the interest payments we will get from the FRA and the payment we could get at no cost, discounted to today of course.

The payment from the FRA will be the notional amount times the 10% rate, or:

$$\text{Payment from FRA} = \$2{,}000{,}000(10\%) = \$200{,}000$$

The payment we could lock in today at *no cost* must also be presented on an annually compounded basis to be able to compare it to the FRA. In this case, since the continuous rate is

9%, the annually compounded rate is:

$$R_F = e^{.09(1)} - 1 = 9.417\%$$

That payment on the same notional would be:

$$\text{Payment at Current Forward Rate} = \$2{,}000{,}000(9.417\%) = \$188{,}349$$

The difference then is $200,000 - 188,349 = \$11{,}651$.

But note that payment would occur at time $T_2 = 3$ years, so the value of the *differences* in payments today is:

$$\text{Value of FRA} = \$11{,}651e^{-.08(3)} = \$9{,}165$$

2007 Exam Question 28

You are given the following loss payout pattern for a property-casualty insurance line of business for a given accident year:

TABLE 1. Payout Pattern

Development Year	% Paid
1	47.20%
2	0.00%
3	0.00%
4	0.00%
5	0.00%
6	0.00%
7	52.80%

- The risk free rate is 4% (annually compounded).
- The yield on the company's investable assets is 9%.
- All loss payouts are made mid-year.
- The environment for cash flow from these insurance losses is non-inflationary.

a. Calculate the liability duration of the loss reserves for this line of business.

Technically, in Feldblum's paper he calculated the duration of loss reserves using a payment pattern for unpaid losses. Here, the wording of the question seems to imply that the pattern given is from inception. But then they ask for the duration of the loss reserves, so it depends entirely on the date — is this at the end of the first development year? If so, then 100% of the unpaid claims are paid 6.5 years from inception and the duration is simply 5.5.

Based on the sample solution, this is NOT what was intended. They intended for the payment pattern to relate to the payment of the outstanding loss reserves and therefore they wanted you to assume that for the current loss reserves, 47.2% would be paid in .5 years and 52.8% would be paid in 6.5 years.

In addition, note that Feldblum recommends using the current yield on investable assets with a similar maturity and risk profile as the liabilities. Given the choices in the problem, the insurer's asset portfolio yield is probably more appropriate than the risk-free rate. Just note that in a more recent paper Feldblum makes it clear that the liability duration is not dependent on the insurer's asset mix and that LIBOR is likely the most appropriate rate to use.

The Macaulay duration can then be calculated as:

TABLE 2. Duration of Loss Reserve

Time to Payment	PV Factor	% Paid	PV(Paid)	T*PV(Paid)
0.5	0.9578	47.20%	0.4521	0.2260
1.5	0.8787	0.00%	0.0000	0.0000
2.5	0.8062	0.00%	0.0000	0.0000
3.5	0.7396	0.00%	0.0000	0.0000
4.5	0.6785	0.00%	0.0000	0.0000
5.5	0.6225	0.00%	0.0000	0.0000
6.5	0.5711	52.80%	0.3016	1.9601
			0.7536	2.1861
			Macaulay Duration:	**2.9007**

Notice that I used the 9% yield on the company's assets for these calculations. In addition, the question did not specify that they wanted the Macaulay duration, but since that's what Feldblum discussed I just calculated that version of the duration measure.

b. For each of the following bond portfolios, demonstrate why or why not the portfolio would be appropriate to duration match the loss reserve for this line of business.

- Portfolio 1 consists of coupon bonds with a coupon of 1% per quarter and a yield of 1.5% per quarter. The bonds mature in 2 years.
- Portfolio 2 consists of level annuities with a 5% yield per year, paid annually, for 5 years.

The first thing to do is calculate the durations of each of these portfolios. The BKM textbook has formulas for these special cases, which you might want to use on the actual exam, but I'll show the brute force way here.

TABLE 3. Duration of Coupon Bond

Time to Payment	PV Factor	CF	PV(CF)	T*PV(CF)
0.25	0.9852	1	0.9852	0.2463
0.50	0.9707	1	0.9707	0.4853
0.75	0.9563	1	0.9563	0.7172
1.00	0.9422	1	0.9422	0.9422
1.25	0.9283	1	0.9283	1.1603
1.50	0.9145	1	0.9145	1.3718
1.75	0.9010	1	0.9010	1.5768
2.00	0.8877	101	89.6588	179.3176
			96.2570	185.8176
			Macaulay Duration:	**1.9304**

TABLE 4. Duration of Annuity

Time to Payment	PV Factor	CF	PV(CF)	T*PV(CF)
1	0.9524	1	0.9524	0.9524
2	0.9070	1	0.9070	1.8141
3	0.8638	1	0.8638	2.5915
4	0.8227	1	0.8227	3.2908
5	0.7835	1	0.7835	3.9176
			4.3295	12.5664
			Macaulay Duration:	**2.9025**

To duration match the liability we want the duration of the assets to equal the duration of the liabilities. This is achieved with Portfolio 2 but not Portfolio 1.

2007 Exam Question 29

You are given the following information about an insurance company:

- Policyholder surplus as of December 31, 2005 = $100
- Premium of policies written January 1, 2006 = $100
- Expected losses on January 1, 2006 policies = $80
- Expenses on January 1, 2006 policies = $25
- Risk free rate = 4%
- 2007 Client retention rate = 95%
- 2008 Client retention rate = 80%
- 2009 Client retention rate = 0%

Assume expenses are paid when the policy is written and losses are certain and paid at the end of the year.

a. Calculate the current economic value of this firm on January 1, 2006.

Panning defined the *current economic value* as the value of the surplus plus the present value of the premiums, expenses and losses on policies already written. Since premiums and expenses are paid immediately, they are not discounted, but the losses are discounted for 1 year.

TABLE 5. Calculating Current Economic Value

Year	Retention	Premiums	Expenses	Losses	PV Premium	PV Expenses	PV Losses	Total PV
2006	100%	100	25	80	100	25	76.92	-1.92
							Surplus:	100
							Value of 2006 Policies:	-1.92
							Current Economic Value:	98.08

b. Calculate the total economic value of this firm on January 1, 2006.

The total economic value adds the present value of future premiums, expenses and losses to the current economic value. The 2006 figures used above to get the current economic value can be extended to include 2007 and 2008. In the Panning paper he assumed business was written in perpetuity, but here the retention rate drops to zero in the third year, so we can do this manually as follows:

TABLE 6. Calculation of Total Economic Value

Year	Retention	Premiums	Expenses	Losses	PV Premium	PV Expenses	PV Losses	Total PV
2006	100%	100	25.00	80.00	100.00	25.00	76.92	-1.92
2007	95%	95	23.75	76.00	91.35	22.84	70.27	-1.76
2008	80%	76	19.00	60.80	70.27	17.57	54.05	-1.35
2009	0%	0	0.00	0.00	0.00	0.00	0.00	0.00
							Franchise Value (2007 and 2008 Policies):	-3.11
							Current Economic Value (from Part a):	98.08
							Total Economic Value:	94.97

Notice that since all of the premiums, expenses and losses just scale by the same factor (the retention rate), we could have used the 2006 present value of -1.92 and just found the 2007 value using the retention rate and one more year of discount, or $-1.92(.95)/1.04 = -1.76$. Similarly, the 2008 value would be $-1.76(.8)/1.04 = -1.35$.

Note that the CAS sample solution contains an error in the formula shown for the franchise value when the retention rate is constant. They didn't use that formula in their solution, so their numbers are right. Just be careful about the erroneous formula.

c. Suppose a rating agency determines that the duration of the total economic value is high and is considering lowering the insurance company's rating. Discuss aspects of the dilemma created if the insurance company uses traditional duration-matching to manage its interest rate risk.

If the company used traditional duration matching to manage the duration of its total economic value, it would have to invest in assets with very low and likely negative durations (e.g. using exotic interest rate derivatives). However, since the balance sheet does not show the franchise value, third parties would likely be given the mistaken impression that they were mismatched and taking on excessive risk (e.g. from the derivatives). So even though they were prudently managing the shareholders' overall interest rate sensitivity properly, it wouldn't appear to be the case. An alternative approach might be to directly manage the interest rate sensitivity of the franchise value itself, which could be done through a pricing strategy that made premiums more responsive to changes in interest rates.

2010 Exam Question 24

Given the following information for a property-casualty insurer as of December 31, 2009:

- Surplus as of December 31, 2009 is $100
- 2010 Expected Loss and LAE is $150
- 2010 Expenses are $50
- The insurer writes all of its business and pays all expenses on Jan each year
- The insurer pays 100% of the losses on December 31 each year
- Surplus is constant each year
- The insurer has no taxes and no risk of bankruptcy
- The client retention is 85% each year
- Premiums, losses and expenses are constant each year for new business, but are impacted by the client retention in all renewal years
- The risk-free interest rate is constant at 5% compounded annually
- The target return on surplus is constant at 15%

a. Calculate the total economic value of the firm as of December 31, 2009.

There are several steps here, so let's follow Panning's approach and handle this in order. I will add an extra step just to review each of Panning's calculations.

First, we need to know how much premium is charged each year. Note that Panning assumes the firm prices its policies such that it expects to earn a return on surplus of k. As a result, the sum of the premiums, losses, expenses and investment income (at a risk-free rate y) on the surplus and net premium must equal k times the initial surplus:

$$P - L - E + (S + P - E)y = kS$$

Solving for P, we have the following formula for the premium that would be charged given these other assumptions:

$$\begin{aligned} P &= \frac{S(k-y) + L}{1+y} + E \\ &= \frac{100(.15 - .05) + 150}{1.05} + 50 \\ &= 202.38 \end{aligned}$$

Next, we need to determine the present value profit of the 2010 business and add that to our current surplus to get the current economic value.

$$\text{Current Economic Value} = \text{Surplus} + \text{Premium} - \text{Expenses} - \text{PV(Losses)}$$

$$\begin{aligned} C &= S + P - E - \frac{L}{1+y} \\ &= 100 + 202.38 - 50 - 150/1.05 \\ &= 109.52. \end{aligned}$$

Notice that I did an extra step than I needed to because the present value profit of 9.52 is just the return over the risk free rate we targeted on the surplus on a present value basis, or:

$$\text{PV(Profit)} = S(k - y)/(1 + y) = 9.52$$

In any case, now we need to know the franchise value. This depends on the steady-state annual profit of 9.52 but adjusted to account for the 85% retention rate and the discounting of each of those future profits.

Using Panning's notation, we define the retention rate as cr and define a discount factor $d = \frac{cr}{1+y} = .8095$. Then, franchise value is simply:

$$\begin{aligned} F &= [\text{PV Profit Current Year}]\left[\frac{d}{1-d}\right] \\ &= 9.52\left[\frac{.8095}{1-.8095}\right] \\ &= 40.48 \end{aligned}$$

And finally, the total economic value is the sum of the current economic value and the franchise value, or \$150.

b. It has been argued that Asset-Liability Managment (ALM), as typically practiced, focuses on the balance sheet assets and liabilities recognized by accounting rules and fails to recognize a significant component of total economic value. Describe the component of total economic value that traditional ALM ignores under this argument.

Traditional ALM ignores the franchise value of the company, which as we saw in the previous question represented almost 30% of the total value of the firm. This portion of the firm value is also sensitive to the effects of changes in interest rates and so if only the current economic value, as depicted on the balance sheet, is managed with respect to interest rate risk, much of the shareholders' value is not being managed in a similar way. An immunized current value could still result in a total shareholder value that is very sensitive to changes in interest rates.

c. Suppose the firm believes that its duration of total economic value is too large. The traditional approach to reducing this duration is to reduce the duration of invested assets by either changing the composition of its investment portfolio or by purchasing derivative

securities that modify its asset duration. Describe two aspects of the practical dilemma created by this approach to reducing the duration of total economic value and describe an alternative approach to reducing the duration of total economic value.

The issue is that it may be difficult to reduce the duration of the investment portfolio low enough to offset the high duration of franchise value without buying exotic-looking securities with very large negative durations. This will then run the risk of making your balance sheet appear as though you are taking exotic interest rate bets imprudently. Secondly, the traditional measurement of asset and liability durations will appear to show a large mismatch, again suggesting a risky strategy.

An alternative is to manage the interest rate sensitivity of the franchise value through an explicit pricing strategy that is itself sensitive to changes in interest rates. This will directly reduce the duration of the franchise value and manage this invisible risk in an invisible way.

2006 Exam Question 32

An insurance company has a liability of $14,500 that will be paid at the end of three years. They are considering funding this liability by purchasing a $10,000 bond that pays 10% annual coupons and matures in six years. All coupons are to be reinvested.

The current yield to maturity is 5.0% with annual compounding.

Determine if this bond is an appropriate choice to fund the obligation and immunize the insurer from small changes in interest rate levels and explain your answer. Ignore taxes in your answer.

To test whether this is an appropriate asset for the liability of $14,500 to be paid in 3 years, one approach would be to test what the accumulated value of the bond's cash flows would be on the date the liability is paid if the coupons are reinvested, at the current yields, and the bond is sold at $T = 3$.

At that time, the coupons would have an accumulated value of $3,153. The bond would have 3 years remaining until maturity and be worth $11,362. The total proceeds available to pay the liability would be $14,514.

These calculations are shown below:

TABLE 7. Future Value of Bond

Coupon	Value at T=3
1,000	1,103
1,000	1,050
1,000	1,000
	3,153
Bond Sale:	11,362
Total:	14,514

As this shows, the bond's cash flows will be sufficient to pay the liability *if rates remain unchanged*. However, if rates were to rise suddenly to, for instance 6%, then the lower value when the bond has to be sold will not be offset by higher values of the reinvested coupons. In this case, the total value available to pay the liability will be only $14,253.

This occurs because the bond's Macaulay duration is 4.9 and the liability horizon is only 3 years. Only when the Macaulay duration is equal to the horizon (3 years in this case) are gains from reinvesting coupons at the higher rate sufficient to offset the loss in value when the bond is sold. That would make this bond inappropriate to immunize the insurer against changes in yields.

2008 Exam Question 31

An insurance company has an obligation to make a payment of $15,485 eight years from today. The market interest rate is 11% annually compounded (assume that there is no credit risk for this insurance company so that the 11% is both the risk free rate and the rate applicable to valuing the insurer's obligation). The insurer plans to fund the obligation using zero coupon bonds that mature in 4 years and perpetuities that pay annual coupons.

Determine the amounts to be invested today by the insurer in perpetuities and zero coupon bonds to immunize the obligation.

Recall that to immunize an obligation you simply need to set the durations of the obligation equal to the duration of the hedging portfolio. It generally doesn't matter whether you use Macaulay or Modified durations, but in this case the Macaulay durations are easier.

Recall that for zero coupon bonds the Macaulay durations are simply equal to the time to maturity. So the duration of the obligation is 8.0 and the duration of the zero coupon bonds used for the hedge is 4.0.

The trickier issue is the Macaulay duration of the perpetuity. The formula is given in the textbook, but it is very easy to derive it by first calculating the Modified duration. If the annual coupon is c then the value of the perpetuity is simply c/y. Taking the derivative with respect to y we have $-c/y^2$. Then recall that to define the modified duration we ignore the negative sign and divide by the value of the bond. This gives a modified duration of $1/y$.

Finally, the relationship between Macaulay and Modified duration is:

$$\text{Modified Duration} = \frac{\text{Macaulay Duration}}{1+y}$$

And so the Macaualay duration of the perpetuity is:

$$\frac{1+y}{y} = \frac{1.11}{.11} = 10.091$$

To get the hedge portfolio to have a duration of 8.0, we would invest w in the zero coupon bond and $(1-w)$ in the perpetuity such that:

$$8 = w(4) + (1-w)(10.091)$$

Then solve for $w = 0.34328$.

But we also want the portfolios to have the same present values. The present value of the obligation is $6,719 and so the present value dollar amounts invested in each asset are $2,306.6 in the zero coupon bond and $4,412.70 in the perpetuity.

2010 Exam Question 22

An insurance company needs to make a payment of $1,967.15$ in ten years. The insurance company would like to fund the obligation using four-year zero coupon bonds and perpetuities paying annual coupons.

The interest rate is 7% compounded annually.

a. Calculate the amount the insurance company should invest in four-year zero coupon bonds and perpetuities paying annual coupons in order to immunize the obligation from interest rate fluctuations.

This is identical to the previous question from the 2008 exam, just with different values.

Our obligation has a Macaualay duration of 10, since it is a zero coupon obligation. Similarly, the four year zero coupon bond has a Macaulay duration of 4. And finally, the perpetuity has a Macaulay duration of $1.07/.07 = 15.286$.

Now we just have to determine a weight, w for the four year bond so that the weighted average duration is equal to 10. We solve $10 = w(4) + (1 - w)15.286$ to get $w = 46.8\%$ invested in the four-year bond.

In addition to matching the durations we also want to match the present values. We know the present value of the obligation is $1,967.15(1.07^{-10}) = 1,000$. Therefore, we need to invest in four year bonds with a current value of \$468 and in perpetuities with a present value of \$532.

2010 Exam Question 7

Given the following information:

Maturity	Price of Zero Coupon Bond
1 year	913.93
2 year	818.73

Six months ago ABC Company entered into a forward rate agreement where it agreed to receive 10% with annual compounding on a principal amount of $10 million between time period $T_1 = 1.5$ years and $T_2 = 2.5 years$. Assume interest rates are compounded continuously.

Calculate the value of this forward rate agreement today.

The agreement was struck 6 months ago, so that means that as of today the agreement involves a 10% interest payment, made in 2 years. If we wanted to sell this agreement to someone else, we need to realize that they can lock in a forward rate between $T = 1$ and $T = 2$ on their own at the current forward rate, which we calculate from the zero coupon rates (9% and 10% respectively based on the bond prices):

$$F = \frac{10\%(2) - 9\%(1)}{2 - 1} = 11\%$$

This is on a continuous basis. Converting it to an annual basis, we get $F_A = e^{.11} - 1 = 11.63\%$.

This means that if we sold it to someone, they would receive 10% interest at $T = 2$ when they could have locked in 11.63% on their own. So to entice them to take this FRA from us, we would have to pay them the difference in the cash flows, on a present value basis. This means the FRA currently has a value to us of -133, 270.

The calcs are shown below:

1 year rate	9.00%
2 year rate	10.00%
Forward rate - continuously compounded	11.00%
Forward rate - annually compounded	11.63%
Notional	10,000,000.00
Agreed Rate (K)	10.00%
Forward Rate	11.63%
Difference in Payments	(162,776.50)
PV Difference	(133,270.00)

2012 Exam Question 2

You are given the following information regarding four $\$1,000$ par value zero coupon bonds and annually compounded forward rates:

Maturity (Yrs)	Price	Year	Forward Rate
1	952.38	1	5.0%
2	902.73	2	5.5%
3	845.55	3	5.5%
4	807.23	4	6.0%

Design an arbitrage strategy and calculate the resultant profit.

First, we need to determine if any of the bonds are mispriced. We assume the forward rates are correct and determine the correct prices for each of the bonds as:

$$\text{Price} = \frac{1,000}{(1 + r_1)(1 + f_2)\ldots(1 + f_n)}$$

These prices are shown below:

Maturity (Yrs)	Price
1	952.38
2	902.73
3	855.67
4	807.23

From this table, we can see that the 3-year bond is underpriced by $10.12. To take advantage of this as an arbitrage, we need to be able to earn a risk-free profit. We can do this by buying the (cheap) 3-year bond for $845.55 using borrowed funds. If we borrow the present value of $1,000 at the current forward rates for 3 years, we can get $855.67 today, out of which we use $845.55 to buy the cheap 3-year bond. At maturity we collect the principal on the cheap bond and use it to repay our borrowing. The net, risk-free profit today is $10.12.

2012 Exam Question 4

You are given the following information regarding a company's loss reserves for accident year 2013:

- Projected loss payout pattern by development year – 2013: 30%, 2014: 35%, 2015: 25%, 2016: 10%
- Losses are paid at the end of each development year
- Inflation rate is projected to be 0% for the next four years
- The risk-free rate is 2%, compounded annually
- The annual yield on the company's investable assets is 4%, compounded annually
- Ultimate losses for accident year 2013 are expected to be $5 million

Calculate the Macaulay duration of loss reserves for accident year 2013 as of December 31, 2013

The payout pattern given reflected the payout from inception, but the question (and the Feldblum paper it is based on) asked about the reserve payout as of the end of the first development year. Taking into account the size of the liability, the following table shows the duration calculation:

T	CF	PV	PV(CF)	T*PV(CF)
1	1.75	0.9615	1.6827	1.683
2	1.25	0.9246	1.1557	2.311
3	0.50	0.8890	0.4445	1.333
Sum			3.2829	5.328
Macaulay Duration				**1.623**

Notice that in the table above I used the annual yield on the company's investable assets rather than the risk-free rate. Feldblum's paper isn't entirely clear on this point and it is common for this to be misinterpreted (including by the exam graders).

In the Feldblum paper on the syllabus, he says specifically that the proper discount rate to use is the rate on *new investable assets*. This would suggest that the current investment portfolio is irrelevant and the only thing that matters is the current market yields.

A more recent ALM paper by Feldblum makes this clearer. In that paper, he says that LIBOR rates are probably better than US Treasury rates, but confirms that the current portfolio yield is irrelevant. He says, "Some analysts use the average yield on the asset portfolio backing the reserves as the discount rate for the reserves...This view is rejected by mainstream investment theory. Just as asset duration is an attribute of the security, not of the other securities held by the investor, the duration of a liability is an attribute of the liability, not of the company holding the liabilities."

Illustrate how to immunize the loss reserve liability of accident year 2013 (as of December 31, 2013) from interest rate fluctuations. Describe two limitations associated with the proposal.

To immunize the liability, we need to hold assets with a present value equal to the present value of the liability (\$3.28m) and the same Macaulay duration (1.623). This will only protect against the effect of small, instantaneous changes in the interest rate and will need to be continually rebalanced to maintain matched durations.

2005 Exam Question 25

ABC Insurance Company has entered into a 10-year currency swap with XYZ Insurance Company. Under the terms of the swap, ABC receives interest at 4% per annum in Swiss francs and pays interest at 8% per annum in US dollars. Interest payments are exchanged once per year. The principal amounts are 10 million US dollars and 13 million Swiss francs.

Suppose that XYZ declares bankruptcy at the end of year 7, right before the swap payment is to be made. The exchange rate at the end of year 7 is \$0.80 per franc.

Assume that forward rates are realized and, at the end of year 7, the interest rate is 4% per annum in Swiss francs and 9% per annum in US dollars for all maturities.

All interest rates are quoted with annual compounding.

Calculate the cost to ABC at the end of year 7 due to XYZ's bankruptcy.

There are four swap payments remaining. If XYZ is bankrupt and fails to meet its remaining obligations, and assuming a zero recovery rate, then the loss to ABC is the remaining value of the swap – if it is positive.

We can value the four remaining payments by assuming that the forward exchange rates will be realized, using those to convert all net payments to dollars and then discounting these at the current risk free rates. These steps are shown below.

The first step is to determine the forward exchange rates. From Hull, the usual formula is:

$$F = Se^{(r-r_f)T}$$

However, the question said that the interest rates given were based on annual compounding. One approach is to convert these to continuous rates and use Hull's formula. Alternatively, we can use the formulas from BKM for the forward rate using annual compounding:

$$F = S\left[\frac{1+r}{1+r_f}\right]^T$$

The forward exchange rates are determined as follows:

TABLE 8. Determining the Forward Rates

Current Exchange Rate	0.80
USD risk free rate	9.00%
Franc risk free rate	4.00%
Maturity	Forward Rate
0	0.8000
1	0.8385
2	0.8788
3	0.9210

Using these rates, the cash flows and their present values are determined as follows:

TABLE 9. Value as a Series of FRAs

	Cash Flows ABC Receives				
Maturity	Notional	Swap Rate	Payment in Francs	Forward Rate	Payment in USD
0	13,000,000	4.00%	520,000	0.8000	416,000
1	13,000,000	4.00%	520,000	0.8385	436,000
2	13,000,000	4.00%	520,000	0.8788	456,962
3	13,000,000	4.00%	13,520,000	0.9210	12,452,202
	Cash Flows ABC Pays				
Maturity	Notional	Swap Rate	Payment in USD		
0	10,000,000	8.00%	800,000		
1	10,000,000	8.00%	800,000		
2	10,000,000	8.00%	800,000		
3	10,000,000	8.00%	10,800,000		
	Cash Flows to ABC				
Maturity	Receives	Pays	Net	PV Cash Flows	
0	416,000	800,000	-384,000	-384,000	
1	436,000	800,000	-364,000	-333,945	
2	456,962	800,000	-343,038	-288,729	
3	12,452,202	10,800,000	1,652,202	1,275,803	
			561,163	**269,129**	

The value of the swap to ABC is \$269,129 when XYZ goes bankrupt, so the cost is \$269,129.

Alternative Solution to 2005 Exam Question 25

Notice that there is another way to calculate this cost, treating ABC's position as being long a franc bond and short a USD bond.

Table 10. Value as Exchange of Bonds

Maturity	Notional	Swap Rate	Coupons	Principal	Bond Value
			Value of Franc Bond		
0	13,000,000	4.00%	520,000	0	520,000
1	13,000,000	4.00%	520,000	0	500,000
2	13,000,000	4.00%	520,000	0	480,769
3	13,000,000	4.00%	520,000	13,000,000	12,019,231
					13,520,000
				Exchange Rate	0.80
				Dollar Value of Franc Bond	10,816,000

Maturity	Notional	Swap Rate	Coupons	Principal	Bond Value
			Value of Dollar Bond		
0	10,000,000	8.00%	800,000	0	800,000
1	10,000,000	8.00%	800,000	0	733,945
2	10,000,000	8.00%	800,000	0	673,344
3	10,000,000	8.00%	800,000	10,000,000	8,339,582
					10,546,871
				Dollar Value of Long Franc Bond	10,816,000
				Dollar Value of Short USD Bond	-10,546,871
				Swap Value	**269,129**

Part 4

Financial Risk Management (a)

Part 4 Table of Contents

Background: Options

The CAS Exam 9 syllabus assumes that all students have been exposed to option pricing using the Black-Scholes formula from either previous attempts at Exam 8 or from the preliminary exams such as Exam 3F. Since some students may not have been exposed to this material, or more likely simply don't recall the details, I am including this brief summary of the key points that are relevant to the other readings on this syllabus.

I will not present a complete review of option pricing here. Instead, I will focus solely on the following:

- Definitions of Put and Call Options
- Put-Call Parity
- Black-Scholes-Merton Option Pricing Formula
- Ito's Lemma

The readings that reference these specific topics include Hull Chapter 23, Kreps' *Investment Equivalent Reinsurance Pricing*, Butsic's Expected Policy Deficit paper and Cummins' *Allocation of Capital in the Insurance Industry*. Please refer to this background summary if the references in those papers are confusing or unclear to you.

Definitions of Put and Call Options

A call option gives the option buyer the right (the option) to BUY an asset (stock, currency, index, etc.) for a given price (the strike price or exercise price) at some date (the expiration date). Because you do not have to buy the asset, you will do so only if it will lead to a profit. Therefore, the intrinsic value of the option on the expiration date will be $\max(S - K, 0)$, where S is the price of the asset and K is the exercise price.

A put option is similar but gives you the option to SELL the asset for the strike price.

An American option can be exercised any time up to the expiration date whereas a European option can only be exercised on the expiration date. The subsequent readings on the Exam 9 syllabus will refer only to European options, so this should make life simple.

Note that you can either be the buyer or the seller of an option. If you sell a call option, then you are said to *short* the call option and you have given someone else the right to buy the asset from you on the expiration date. You do not have a say in the matter — only they do — and their gain is your loss. So if at expiration $S > K$, they will exercise and you will lose $K - S$ (ignoring the money they paid you for the option). Otherwise, they will not exercise and nothing will happen — you will just keep the money they paid you up front for the option.

Option Values

The following are the key factors that affect the value of an option:

- Current Stock Price — The intrinsic value of the option is directly related to the price of the underlying asset, so this should be obvious. Call options increase in value as the stock price rises; put options decrease as the stock price rises.

- Exercise Price or Strike Price — For a call option, the exercise price specifies how much you have to pay to obtain the stock, so clearly the option value declines as this rises. Similarly, for a put option the value will rise as the exercise price rises, since it reflects the price at which you can sell the stock.

- Time to Expiration — All options essentially get the bulk of their value from the fact that prices may rise or fall prior to expiration. The longer this time period, the more valuable the option. However, because of complications from the effects of dividends, it may be the case that long dated European options are less valuable than short dated European options because in this case you are limited to exercise only at expiration and dividends paid in the interim could hurt you.

- Volatility — This is a measure of how much the stock price changes during a period of time. It is related to the standard deviation of the stock returns. Because call option holders benefit from price increases but don't suffer comparable losses from price decreases (payoffs cannot be negative), increased volatility increases the value of an option. The same is true for put options.

- Risk Free Rate — This is a bit tricky. As rates rise, two things happen. First, the expected growth rate of the stock increases, which will tend to make calls more valuable and puts less valuable. Second, the higher rates means the present value of the option payoffs are lower. For calls, the effects are offsetting, with the first effect dominating so that as rates rise, call values rise. For puts, both effects decrease the value.

- Dividends — Since dividends decrease the stock price, the effect on option values is the same as mentioned above for stock price declines. In what follows, I will generally ignore the effect of dividends since the subsequent readings on the 2011 Exam 9 syllabus don't reference the cases where dividends (or similar cash flows) exists.

Notation

The following notation will be used for these notes, which may differ slightly from the notation used when you learned this material before.

- S = current stock price

- S_T = stock price at time T

- K = strike price
- σ = volatility
- r = risk free rate
- T = time to expiration (in years)
- c = value of European call
- p = value of European put

Put-Call Parity

This is an extremely important relationship. You should understand the derivation and be able to apply this relationship in numerical problems.

Consider European puts and calls on a stock that does not pay dividends. We will form two different portfolios and show that because they have the same final payoffs at maturity regardless of what happens to the stock price, they must have the same initial value.

Portfolio A contains one European call and cash equal to Ke^{-rT}. Portfolio B contains one European put and one share of the stock. Both the put and the call have strike prices of K. Table 1 shows the possible payoffs at time T of each portfolio.

TABLE 1. Payoffs at Expiration

		$S_T < K$	$S_T \geq K$
Portfolio A:	European Call	0	$S_T - K$
	Cash	K	K
	Total	K	S_T
Portfolio B:	European Put	$K - S_T$	0
	Stock	S_T	S_T
	Total	K	S_T

Notice that these two different portfolios have identical payoffs in both possible states of the world. In an efficient market, two assets with identical payoffs must have the same price. Therefore, the cost of buying Portfolio A must equal the cost of buying Portfolio B, which leads to an important relationship between the prices of puts, calls, the underlying stock and the risk free rate known as **Put-Call Parity**:

$$c + Ke^{-rT} = p + S$$

This tells you what the relationship must be between the price of a put and the price of a call to avoid an opportunity for arbitrage.

The Black-Scholes Formula

Without worrying at all about the derivation for a moment, it is important for the subsequent readings that students are familiar with the basic form of the Black-Scholes option pricing formula for a *European call option*.

$$c = SN(d_1) - Ke^{-rT}N(d_2)$$

where $N(\cdot)$ is the standard normal CDF, and d_1 and d_2 are defined as follows:

$$d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S/K) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

Derivation

There are several ways to derive this famous result, but for our purposes we simply need to be aware of one key issue — *Risk-Neutral Valuation*.

It would be relatively easy, for a given model of the stock price distribution at maturity, to derive a formula for the expected payoffs from a call option. However, we wouldn't be able to determine the current value of this option because these cash flows would be risky and so we would need to know an appropriate risk-adjusted discount rate to use. Without any guidance, we are left with the ability to estimate expected cash flows, but not assign a *value* to them.

It turns out (refer to Hull Chapters 11-13 for details) that we can get around this problem using a trick called risk-neutral valuation. We simply assume that investors are risk-neutral such that they value all cash flows using risk-free discount rates. A subtle consequence of this assumption is that it implies a different distribution for the underlying stock price than we might believe applies in the "real world", so we need to take this into account when we determine the (risk neutral) expected cash flows from the option.

To do this, we know the payoff to a call option is $\max(S_T - K, 0)$. Since S_T is a random variable, the math to determine the expected value of this function is messy but merely involves an integral.

It is interesting to note that the math for this is no different than the formula for the expected excess insurance losses for a lognormal loss function as covered in the preliminary CAS exams. The key result is that for any lognormal random variable, V, with parameters m and s (i.e. the mean of $\ln(V)$ is m and the standard deviation is s):

$$\begin{aligned} E[\max(V_T - K, 0)] &= \int_K^{\infty} (V - K) g(V)\, dV \\ &= E(V_T) N(d_1) - K N(d_2) \end{aligned}$$

where,

$$\begin{aligned} d_1 &= \frac{\ln[E(V_T)/K] + \frac{1}{2}s^2}{s} \\ d_2 &= \frac{\ln[E(V_T)/K] - \frac{1}{2}s^2}{s} \end{aligned}$$

Deriving the Black-Scholes formula from this result is relatively trivial. The only tricky part of this is that we have to assume that nobody cares about risk. This means that **all** assets earn the risk free rate of return, including the stock itself. As a result, we need to adjust our probability distribution for the stock price such that the expected return equals r rather than μ.

Once we do this, the expected value of the option payoff is pretty easy and then the call is simply the present value of that amount — also discounted at the risk free rate.

This merely involves setting $V = S$, plugging in for $E(S_T)$ given the assumption that the expected return on S is the risk free rate (i.e. $E(S_T) = Se^{rT}$) and plugging in $s = \sigma\sqrt{T}$. Then,

$$\begin{aligned} c &= e^{-rT} E[\max(S_T - K, 0)] \\ &= e^{-rT}[Se^{rT} N(d_1) - K N(d_2)] \\ &= SN(d_1) - Ke^{-rT} N(d_2) \end{aligned}$$

Where $N(\cdot)$ is the standard normal CDF, and d_1 and d_2 are defined as follows:

$$\begin{aligned} d_1 &= \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \\ d_2 &= \frac{\ln(S/K) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \\ &= d_1 - \sigma\sqrt{T} \end{aligned}$$

Black-Scholes Formula for European Put

While it is easy to use the same approach as above to derive a formula for a put option, in practice it is almost always easier to value a put by first valuing the call option with the same strike price and maturity date and the use Put-Call Parity to get the put value. However, in certain subsequent readings on the 2011 CAS Exam 9 syllabus the authors make reference to the Black-Scholes Put formula, so it is helpful to see what that looks like.

We'll start with Put-Call Parity to get a formula for the put option, plug in the Black-Scholes formula for the call option and then simplify:

$$\begin{aligned} c + Ke^{-rT} &= p + S \\ p &= c + Ke^{-rT} - S \\ &= [SN(d_1) - Ke^{-rT}N(d_2)] + Ke^{-rT} - S \\ &= S[N(d_1) - 1] - Ke^{-rT}[N(d_2) - 1] \\ &= Ke^{-rT}[1 - N(d_2)] - S[1 - N(d_1)] \\ &= Ke^{-rT}N(-d_2) - SN(-d_1) \end{aligned}$$

Special Case: Strike Price Equals the Forward Price

An interesting special case, referenced by Kreps in his paper on the syllabus, is when the strike price is equal to the *forward price* of the underlying asset, or $K = Se^{rT}$. This strike price represents that value the stock price would grow to if it earned the risk-free rate of return.

Using Put-Call Parity, we see that in this special case the value of the call and the value of the put are identical:

$$\begin{aligned} c + Ke^{-rT} &= p + S \\ c + (Se^{rT})e^{-rT} &= p + S \\ c + S &= p + S \\ c &= p \end{aligned}$$

It is also interesting to note what the Black-Scholes formula simplifies to in this special case:

$$\begin{aligned} c &= SN(d_1) - Ke^{-rT}N(d_2) \\ &= SN(d_1) - (Se^{rT})e^{-rT}N(d_2) \\ &= S[N(d_1) - N(d_2)] \end{aligned}$$

where

$$\begin{aligned} d_1 &= \frac{\ln\left(\frac{S}{K}\right) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \\ &= \frac{\ln\left(\frac{Se^{rT}}{K}\right) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \\ &= \frac{\ln\left(\frac{K}{K}\right) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \\ &= \frac{1}{2}\sigma\sqrt{t} \\ d_2 &= -\frac{1}{2}\sigma\sqrt{t} \end{aligned}$$

This makes the formula for a put or a call in this case rather simple:

$$\text{Option Value} = S\left[N\left(\frac{1}{2}\sigma\sqrt{t}\right) - N\left(-\frac{1}{2}\sigma\sqrt{t}\right)\right]$$

And just for fun, notice that there is a pretty decent approximation for this formula, which is:

$$\begin{aligned} c &\approx \frac{1}{\sqrt{2\pi}} S\sigma\sqrt{T}\left(1 - \frac{\sigma^2 T}{24}\right) \\ &\approx .4S\sigma\sqrt{T} \end{aligned}$$

Kreps references this approximation in his paper on the syllabus, so it could perhaps be used as a calculation shortcut on the exam because it avoids having to look up values in the normal distribution table.

Ito's Lemma

It should be noted that the approach outlined above is not actually the basis for the original Black-Scholes paper's derivation of their famous formula.

In one of their two derivations, based on insights from Robert Merton, they showed that one could combine the call option with the underlying stock, in some proportion, to produce an instantaneously riskless portfolio. Using a particular model for the stochastic process followed by the stock, they were able to derive a formula for the stochastic process followed by this hedged portfolio using what is known as Ito's Lemma and showed that this portfolio is in fact risk-free (there is no volatility). Finally, they used the argument that a risk-free portfolio ought to earn the risk free rate of return in order to derive a formula for the call option.

I won't show the derivation here, but you do need to be familiar with the stock price process they assumed and Ito's Lemma.

Stock Price Process

The model used for the stock price is known as *Geometric Brownian Motion.* The formula is as follows:

$$dS = \mu S\, dt + \sigma S\, dz$$

This looks complex, but it can be explained quite easily:

- The current stock price is S and over some period of time, Δt, the price can change by some amount, ΔS. If we assume the time period is very short, then we can use the notation dt and dS to denote the change in time and the change in the stock price.

- The amount of the stock price change, dS, has two components.

  a. The first component, $\mu S\, dt$, is just the expected amount by which the price changes per unit of time and consists of the expected return of μ times the stock price times the amount of time dt.

  b. The second component, $\sigma S\, dz$, is a *random* component. The size of this depends on the volatility of the stock price, σ times the stock price times a normally distributed random variable dz.

- Focus a bit on the random component, dz. All you need to know about this is that it is normally distributed with a mean of zero and a variance equal to dt. When we use a small interval of time Δt, then we can write this random component as the product of a standard normal component ϵ with mean zero and variance 1, and a time component $\sqrt{\Delta t}$. So the random component is given by $dz = \epsilon\sqrt{\Delta t}$.

Ito's Lemma

Recall that our goal is to understand the stochastic process followed by a combination of a call option and some number of shares of the underlying stock. We know, or assumed, the nature of the stock price process, so now we just need to know the process for the option value.

Because the option value is a function of the stock price and time (among other things), according to the formula:

$$dS = \mu S\, dt + \sigma S\, dz$$

we know from first year calculus that the change in the option value, call it dG, is related to dS and dt. If dS did not have a random component, then we could use our basic calculus results to find this derivative involving partial derivatives of both S and t.

But because of the random component, and the fact that dS itself is a function of dt, the math is a lot messier and the resulting formula for dG is somewhat intimidating.

Recall from calculus that if $G(S,t)$ is a function of S and t, then dG can be approximated using a Taylor expansion (see the notes for BKM Chapter 16) as:

$$dG = \frac{\partial G}{\partial S} dS + \frac{\partial G}{\partial t} dt + \frac{1}{2}\frac{\partial^2 G}{\partial S^2} dS^2 + \frac{1}{2}\frac{\partial^2 G}{\partial t^2} dt^2 + \frac{\partial^2 G}{\partial S \partial t} dS\, dt + \cdots$$

This formula is very similar to the Taylor expansion we used in the discussion of duration and convexity in the BKM Chapter 16 notes, but with terms related to the first (partial) derivative with respect to *each* variable (S and t), terms related to the second (partial) derivative with respect to *each* variable and a term related to the derivative with respect to both variables.

Of course, there are additional terms involving the third partial derivatives, fourth partial derivatives, etc. However, because by definition dt is a very small number, anything with a $(dt)^2$ is a tiny number so all of those terms can be ignored. In fact, anything with $(dt)^{1.5}$ can also be ignored and we'll only worry about terms that are of order dt.

We already made the assumption that the stock price followed:

$$dS = \mu S\, dt + \sigma S\, dz \ = \mu S\, dt + \sigma S\, \epsilon\sqrt{dt}$$

So we just plug that in to the formula shown above. But notice something. There's a $(dS)^2$ term in the formula for dG, and when we square dS, we get a term with dt^2, a term with $dt^{1.5}$ and a term with dt. As before, dt is small, so anything with dt to any order higher than 1 can be ignored.

And for the same reason, the fifth term in the above Taylor expansion has a $dS\, dt$ term, which results in several terms that are of the order $(dt)^2$ and $(dt)^{1.5}$, so this too can be ignored.

The resulting formula for dG is then a mess, but easy to write. We just plug in for dS and ignore anything with a dt term of higher order than 1. When we do that, we get:

$$dG = \frac{\partial G}{\partial S}(\mu S\, dt + \sigma S\, dz) + \frac{\partial G}{\partial t} dt + \frac{1}{2}\frac{\partial^2 G}{\partial S^2}(\sigma^2 S^2 \epsilon^2\, dt)$$

Notice that I am just ignoring the fourth and higher terms in the formula for dG and also ignoring the first two terms in the $(dS)^2$ expansion.

Now just note that we can assume that $\epsilon^2 \approx 1$ (there is some hand waving here — it has a mean of 1 and a very tiny variance) and $dz = \epsilon\sqrt{dt}$ and then simplify it by multiplying out the terms and factoring the dt and dz terms. The result is known as **Ito's Lemma**:

$$dG = \left(\frac{\partial G}{\partial S}\mu S + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial S^2}\sigma^2 S^2\right) dt + \frac{\partial G}{\partial S}\sigma S\, dz$$

The relevance of this is simply that there is a known relationship between the stochastic process followed by an option and the stochastic process followed by the underlying stock. As shown above, the option value has a more complicated drift term but, importantly, it has a very similar volatility term. In fact, the volatility of the option value is simply the volatility of the stock value, $\sigma S dz$, multiplied by a constant known as the option delta, or $\frac{\partial G}{\partial S}$.

This last point is what made it possible for Black, Scholes and Merton to argue that they could create a riskless portfolio. By holding one call option and *negative* $\frac{\partial G}{\partial S}$ shares of the stock, the volatility terms cancel and the portfolio is riskless.

For our purposes though, we just need to be familiar with the use of Ito's Lemma to show the relationship between the stock volatility and the option volatility, as this will come into play in Hull Chapter 23 when we discuss the Merton Model for default risk.

Practice Questions

Question 1. Assume the current price of a 1-year European call option on Microsoft stock with a strike price of \$110 is \$5. If the current stock price is \$90 and the risk free rate is 10%, what is the value of a 1-year European put on Microsoft with a strike price of \$110?

Solution. This is just a basic application of Put-Call Parity:

$$c + Ke^{-rT} = p + S$$

From this, $p = \$5 + \$110e^{-.1} - \$90 = \$14.53$.

Question 2. The current value of Cisco's stock is \$70 and we will assume that in one year it will either be worth \$80 or \$60 — no other outcomes are possible. Assume the continuously compounded risk free rate is 10% and that 1-year European puts and calls with strike prices of \$75 are available. Determine the possible payoffs to the following two portfolios, A and B, and demonstrate that the payoffs are identical. Assume portfolio A contains a call option and cash equal to the present value of \$75, or \$67.86. Portfolio B contains a put option and one share of the stock. Assume the cash is invested and earns interest at the risk free rate.

Solution. The chart below shows the possible outcomes, depending on the price of the stock in 1 year. The call is worth $\max(S_T - K, 0)$ and the cash grows with interest and will have the same value (\$75) in 1 year regardless of what the stock price turns out to be. The put is worth $\max(K - S_T, 0)$ and the stock will be worth S_T.

TABLE 2. Cash Flows at Expiration

		S_T = \$80	S_T = \$60
A:	Call	\$5	0
	Cash	\$75	\$75
	Total	\$80	\$75
B:	Put	0	\$15
	Stock	\$80	\$60
	Total	\$80	\$75

Notice that the payoffs are either \$80 or \$75 in both Portfolio A and Portfolio B. This means that the value of the two portfolios must be identical and that proves Put-Call Parity:

$$c + Ke^{-rT} = p + S$$

Question 3. What is the value of a 3-month European call option with a strike price of \$40 if the current stock price is \$35, the volatility is 40% per annum and the risk free rate is 6%?

Solution. This is just a basic application of the Black-Scholes formula. The formula is as follows:

$$c = SN(d_1) - Ke^{-rT}N(d_2)$$

Where $N(\cdot)$ is the standard normal CDF, and d_1 and d_2 are defined as follows:

$$d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S/K) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

Plugging in the information given, we find that:

$$d_1 = \frac{\ln(35/40) + (.06 + (.5).4^2)(.25)}{.4\sqrt{.25}} = -0.492$$

$$d_2 = -0.492 - .4\sqrt{.25} = -0.692$$

Now, plug those two amounts into the standard normal CDF (use the Standard Normal Distribution Table in the Appendix and round d_1 and d_2 to 2 decimal points rather than interpolate). Doing this, you will get $N(d_1) = .3121$ and $N(d_2) = .2451$.

Then, plug those values into the Black-Scholes formula:

$$\begin{aligned} c &= SN(d_1) - Ke^{-rT}N(d_2) \\ &= (35)(.3121) - 40e^{-.06(.25)}(.2451) \\ &= 1.26 \end{aligned}$$

Question 4. What is the value of a 3-month European put with an exercise price of \$40 on the same stock described in the previous question?

Solution. The standard Black-Scholes formula is only applicable to valuing call options on stocks that do not pay dividends. To value a put, you need to first use the value of a call, found above to be 1.269, and then use Put-Call Parity, as follows:

$$c + Ke^{-rT} = p + S \Rightarrow p = 1.26 + 40e^{-.06(.25)} - 35 = 5.67$$

Question 5. The Hull text describes that you can rewrite the Black-Scholes formula as:

$$c = e^{-rT}[Se^{rT}N(d_1) - KN(d_2)]$$

and then interpret the term inside the brackets as:

i. the expected value of a variable equal to S_T if $S_T > K$ and zero otherwise, less

ii. the strike price, K, times the probability the strike price will be paid (i.e. the probability that the option will be exercised).

Assume that the current price of Cisco's stock is \$65 and that the stock follows the following stochastic process:

$$dS = .15S\,dt + .3S\,dz$$

Determine what the formula above would estimate as the "probability that a European call option would be exercised" if the exercise price is \$75, the option has 1-year to maturity and the risk free rate is 6%. Determine this probability directly using the stochastic process given for the stock and discuss the differences, if any, between these two estimates.

Remember that if $dS = \mu S\,dt + \sigma S\,dz$ then this tells us that $\ln(S_T)$ is normally distributed with a mean of $\ln(S_0) + (\mu - \frac{1}{2}\sigma^2)T$ and a standard deviation of $\sigma\sqrt{T}$.

Solution. This is a somewhat convoluted but very important question.

The option will be exercised if the stock price is greater than the strike price of \$75, so determining the probability of exercise is the same as determining the probability that the stock price in one year will exceed \$75. The comment in the text indicates that $N(d_2)$ tells you the probability that the option will be exercised in a risk-neutral world. So let's just determine that amount.

Recall the formula for d_2 was:

$$d_2 = \frac{\ln(S/K) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

and notice that all of the information we need is given in the question.

The only thing to be careful about is that I didn't state explicitly what σ was. However, you should have noticed that the stochastic process for the stock price is generally assumed to be $dS = \mu S\,dt + \sigma S\,dz$, so from the information given we know that $\sigma = .30$.

Plugging in to the formula for d_2 gives us:

$$d_2 = \frac{\ln(65/75) + (.06 - (.5).3^2)(1)}{.3\sqrt{1}} = -.427$$

Using the Normal CDF tables, $N(d_2) = N(-.427) = .3347$.

That should have been easy. Now let's use the stock process information directly. We know that when a stock follows this sort of process, the variable $\ln(S_T)$ is Normally distributed with the following mean and standard deviation:

$$\begin{aligned}
\text{Mean} &= \ln(S_0) + (\mu - \frac{1}{2}\sigma^2)T \\
&= \ln(65) + (.15 - .5(.3^2))(1) \\
&= 4.2794 \\
\text{Std Dev} &= \sigma\sqrt{T} \\
&= .3
\end{aligned}$$

To get the probability, note that we want the probability that $S_T > 75$ or $\ln(S_T) > 4.3175$. This is easy for a Normal random variable, and is simply:

$$1 - N\left[\frac{4.3175 - 4.2794}{.3}\right] = 1 - N(.127) = .4495$$

Wait! Why are these two numbers so different?

The first method found the probability to be .3347 and the second method found the probability to be .4495. The key is in the subtle phrasing in the text. The formula discussed in the text indicates that $N(d_2)$ tells you the probability that the option will be exercised ***in a risk-neutral world***. It is not the *real* probability that the stock price will be above \$75, which is what we calculated in the second step as .4495. The $N(d_2)$ term is only the probability in our assumed world where everything earns the risk free rate.

Therefore, to get this same answer that the textbook suggests, we need to modify the stock price process so that the drift rate is the risk free rate of $r = .06$, not the $\mu = .15$ that we used. If we redo the second part, we can see that *in a risk neutral word* the variable $\ln(S_T)$ is Normal with the following mean and standard deviation:

$$\begin{aligned}
\text{Mean} &= \ln(S_0) + (r - \frac{1}{2}\sigma^2)T \\
&= \ln(65) + (.06 - .5(.3^2))(1) \\
&= 4.1894 \\
\text{Std Dev} &= \sigma\sqrt{T} \\
&= .3
\end{aligned}$$

To calculate the probability, note that we again want the probability that $S_T > 75$ or $\ln(S_T) > 4.3175$. This is easy for a Normal random variable, and is simply:

$$1 - N\left[\frac{4.3175 - 4.1894}{.3}\right] = 1 - N(.427) = .3347$$

Notice that this is exactly the value of $N(d_2)$.

Notice that any statements about probabilities when it comes to options and the Black-Scholes formulas are statements about risk-neutral probabilities, not the actual probabilities. That turns out to be a good thing, because if option values actually depended on real probabilities, how would anyone ever be able to price an option? In the real world, nobody knows those probabilities.

Question 6. Suppose you created a portfolio that contained a certain number of shares of stock and a short position in an option on the stock. The option has a value of f. Further assume that the amount of shares you own is equal to the quantity $\frac{\partial f}{\partial S}$ and that the stock price follows Geometric Brownian Motion.

Use Ito's Lemma to determine the stochastic process for the total portfolio value and discuss the degree of riskiness in this process.

Solution. Assume that we have a portfolio containing a short position in the derivative and $\frac{\partial f}{\partial S}$ shares of the stock. The value of the portfolio today would be:

$$\Pi = -f + \frac{\partial f}{\partial S}S$$

To determine the stochastic process that describes the change in the value of this portfolio, simply note that $dS = \mu S\,dt + \sigma S\,dz$ and we can use Ito's Lemma to determine the process followed by the derivative. Using the discrete version of each of these formulas for dS and df:

$$\begin{aligned}
\Delta S &= \mu S\,\Delta t + \sigma S\,\Delta z \\
\Delta f &= \left(\frac{\partial f}{\partial S}\mu S + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)\Delta t + \frac{\partial f}{\partial S}\sigma S\,\Delta z
\end{aligned}$$

Then just plug each of these in to the formula for the change in value of the entire portfolio:

$$
\begin{aligned}
\Delta\Pi &= -\Delta f + \frac{\partial f}{\partial S}\Delta S \\
&= -\left\{\left(\frac{\partial f}{\partial S}\mu S + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)\Delta t + \frac{\partial f}{\partial S}\sigma S\,\Delta z\right\} + \frac{\partial f}{\partial S}(\mu S\,\Delta t + \sigma S\,\Delta z)
\end{aligned}
$$

Simplifying this expression we see that several of the terms cancel out and the result is:

$$
\Delta\Pi = \left(-\frac{\partial f}{\partial t} - \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)\Delta t
$$

Notice the key insight here. The resulting change in the value of the entire portfolio does not have a stochastic term (Δz) and therefore the portfolio has no risk.

Hull Chapter 23: Default Probabilities

This chapter begins with three methods for estimating default probabilities for corporate bonds, using either Historical Data, Bond Prices or Equity Prices (via the Merton Model). Hull then discusses credit risk in derivatives transactions. Finally, he discusses ways to quantify default risk using a Credit Value at Risk measure.

These three topics can be thought of as relatively distinct and so will presented separately in these notes.

In this section of the notes, I will cover only the material related to default probabilities in Sections 23.1 - 23.6 of Hull's 8th Edition. I will address the credit risk in derivatives transactions immediately following. The Credit Value at Risk material will be covered in the next section of the notes.

Background: Default Risk

Other than US government bonds, most bonds have some risk of not making their interest or principal payments. When we value a bond then, we need to discount the expected cash flows taking into consideration this probability of less than full payment.

For instance, suppose a bond promised to pay 100 in one year and the risk-free yield is 10% on an annually compounded basis. If there were no default risk, the price would be:

$$P = \frac{100}{1.10} = 90.9091$$

But, if there's a 5% chance of default and if we assume that in the event of default no payments are made, the bond is only worth the discounted value of its expected cash flows:

$$P = \frac{0.95(100) + 0.05(0)}{1.1} = 86.3636$$

Comparing the price to the promised cash flows, we can see that the *promised yield* is found by setting the price equal to the promised cash flows and solving for the yield:

$$86.3636 = \frac{100}{1+y} \Rightarrow y = 15.789\%$$

Comparing that to the default free yield, we see that there is an 5.789% default risk premium. That is, the promised yield is 5.789% higher than the risk-free yield (both on an annual basis).

However, if the firm does not default and the full 100 is paid, the realized return will in fact be 15.789%, but if the firm defaults the realized return will be −100%.

The *expected yield* is therefore:

$$\text{Expected Yield} = .95(15.789\%) + .05(-100\%) = 10\%$$

Note that this is the same as the expected yield on the default free bond.

Estimates of default probabilities are rather difficult to make, but there are bond rating agencies such as S&P and Moody's that do this. The rating agencies don't actually publish probabilities of default though. They instead assign a letter rating to each bond that serves as a measure of the relative default probability.

Ratings at or above a certain threshold (BBB) are considered *investment grade* and those that are rated lower are considered *speculative.* Bonds rated below investment grade are also often referred to as *high-yield bonds* or *junk bonds*, indicating simply that they have a high probability of default and a correspondingly high promised yield.

Historical Default Probabilities

One way to estimate default probabilities for bonds of different ratings is by collecting historical data for various cohorts of bonds that start with initial credit ratings and observe the rate of default in each cohort through time.

Table 1 is based on Moody's January 2004 report, *Default and Recovery Rates of Corporate Bond Issuers*. It is identical to Table 23.1 in the text.

TABLE 1. Average Cumulative Default Rates (%), 1970-2003

	1	2	3	4	5	10	20
Aaa	0.00	0.00	0.00	0.04	0.12	0.62	1.55
Aa	0.02	0.03	0.06	0.15	0.24	0.68	2.70
A	0.02	0.09	0.23	0.38	0.54	1.59	5.24
Baa	0.20	0.57	1.03	1.62	2.16	5.10	12.59
Ba	1.26	3.48	6.00	8.59	11.17	21.01	38.56
B	6.21	13.76	20.65	26.66	31.99	50.02	60.73
Caa-C	23.65	37.20	48.02	55.56	60.83	77.91	80.23

From Table 1 we can define several types of default probabilities:

- Cumulative Default Probability
- Unconditional Default Probability
- Hazard Rate/Default Intensity/Conditional Default Probability
- Average Default Intensity

Cumulative Default Probability

This is the amount shown in Table 1 on the facing page. It represents the probability of default prior to year n, as of the time of issuance.

Unconditional Default Probability

Table 1 on the preceding page shows the cumulative probability of default at the time of issuance. For example, at issuance a B-rated bond has a 20.65% probability of defaulting by the end of year three. If we subtract the probability of the bond defaulting by the end of year two, then we can estimate the unconditional probability of default *in year three* of 20.65% − 13.76% = 6.89%.

Hazard Rate/Default Intensity/Conditional Default Probability

An alternative measure is the probability of default in year three *conditional* on the bond surviving until the beginning of year three. To calculate this for the B-rated bond above, note that the probability of surviving until the start of year three is 100% − 13.76% = 86.24%. Therefore, the probability of default in year three conditional on surviving through year two is 6.89%/86.24% = 7.99%.

The 7.99% above is the conditional probability of default for a 1-year period. If we shorten the length of the time period to Δt, then the *instantaneous conditional default probability* can be written as $\lambda(t)\Delta t$ where $\lambda(t)$ is the hazard rate or the default intensity.

Average Hazard Rate or Average Default Intensity

The cumulative default probability until time period t, $Q(t)$, can be written in terms of the average instantaneous default intensity as follows:

$$Q(t) = 1 - e^{-\bar{\lambda}(t)t}$$

For example, based on historical data the probability of an A-rated company defaulting by year 7 is 1.441%. We can quote this in terms of an average 7-year hazard rate using the formula above:

$$\begin{aligned} 1.441\% &= 1 - e^{-\bar{\lambda}(7)7} \\ \bar{\lambda}(7) &= -\frac{1}{7} ln[1 - 1.441\%] \\ &= 0.21\% \end{aligned}$$

Recovery Rates

When a bond defaults, it is not always the case that none of its debts are paid. In most cases there is some small recovery rate. If we define the recovery rate for a bond as the ratio of the market value just after default to the face value of the debt, then we can see that historically the recovery rates have varied based on the seniority of the debt. Table 2 shows recovery rate data for corporate bonds from 1982-2003.

Notice that the recovery rate data shows that the recovery rate, at least historically, depended critically upon the seniority of the debt and whether it was secured by some form of collateral

TABLE 2. Recovery Rate as Percent of Face Value 1982-2003 (Moody's)

Class	Average Recovery %
Sr. Secured	54.44
Sr. Unsecured	38.39
Sr. Subordinated	32.85
Subordinated	31.61
Jr. Subordinated	24.47

or not. But it is also the case that recovery rates are negatively correlated with default rates. That is, in a year with high levels of default the recovery rates are generally lower too.

Estimating the Probability of Default from Bond Spreads

When bonds are issued, the prices and yields on the bonds will reflect the expected loss from defaults, with investors paying more for a bond from a high credit quality issuer than from a low credit quality issuer. The difference in the prices of otherwise comparable risk free bonds and bonds with credit risk reflects the present value expected loss from default.

A simple *approximate* formula can be used to estimate the average annual hazard rate over the life of the bond, $\bar{\lambda}$.

First, assume that the *spread* on a bond, which is the bond yield in excess of the yield on a similar risk-free bond, represents the expected losses each year due to defaults (in fact, it may also reflect other factors such as liquidity). We can denote the spread as s per annum.

Next, the average annual loss from defaults can be estimated as the average annual probability of default each year, denoted $\bar{\lambda}$, multiplied by the loss given default, or one minus the recovery rate, $1 - R$.

If the spread is intended to compensate investors for the loss from defaults, then:

$$s = \bar{\lambda}(1 - R)$$

This can be rearranged to write the annual hazard rate as a function of the yield spread and the recovery rate:

$$\bar{\lambda} = \frac{s}{1 - R}$$

Estimating the Probability of Default from Bond Prices

A more precise approach uses the price of a bond with default risk and the theoretical price of an identical bond (same maturity and coupons) without default risk. These two prices will differ by the amount by which the buyer of the risky bond demands to be compensated for assuming this default risk. In other words, the difference in the prices reflects the present value of the expected default losses.

To solve for the default probability, we can separately calculate the present value of the expected default losses as the product of the probability of default and the *loss given default*, or LGD, and summing across all possible points in time where default can occur. This amount, which is a function of the probability of default, can be set equal to the present value of the expected default losses calculated above using the differences in the bond prices to solve for the probability of default.

Simple Example

Consider a simplified example. Assume we have a $100 face value bond that pays 8% coupons annually and has one year until maturity. Assume the yield on the bond is 6.5% and the risk-free rate is 5% (both continuously compounded). The value of the bond can be found to be:

$$B = 108e^{-.065(1)} = 101.20$$

If the bond were risk-free, it's value would be:

$$G = 108e^{-.05(1)} = 102.73$$

The difference, or $1.53, reflects the present value of the expected default losses.

The other way to calculate the present value of the expected default losses is to first calculate the loss given default. If we assume the recovery rate is 40%, then in the event of a default on the maturity date, we would have expected to receive $108 but instead receive only 40% of the face value, making the loss given default $LGD = 108 - 40 = 68$.

If we denote the probability of default in year 1 as Q, then the following is the present value of the expected default losses, which we can set equal to $1.53:

$$1.53 = Q68e^{-.05(1)}$$

Solving for the probability of default in year 1, we get $Q = 2.365\%$.

Of course, for a bond with longer time to maturity, we would have more points in time when default could occur and so we would need to know the relationship between the probability of default at each point in time and sum over all possible default times. In the Hull chapter, he makes a simplifying assumption that the unconditional probability of default each period is the same and then just has one unknown to solve for. On old exam questions, they made different assumptions, such as that the unconditional default probability in year 2 is twice the default probability in year 1.

For students who may be using Hull's 9th Edition, note that when he presents this material he doesn't use an annual default probability and instead uses an average instantaneous hazard rate. For instance, if the average instantaneous hazard rate in year 1 is $\bar{\lambda}_1$, then the following is the present value of the expected default losses, which we can set equal to $1.53:

$$1.53 = (1 - e^{-\bar{\lambda}_1(1)})68e^{-.05(1)}$$

Solving for the average hazard rate in year 1, we get $\bar{\lambda}_1 = 2.394\%$.

Forward Bond Prices

The above example glossed over one technicality that is addressed in the textbook's presentation of this material. When you calculate the loss given default, the relevant value for the bond *at the moment of default* is the *forward risk-free price* of the bond. That amount is calculated as the present value of the remaining cash flows on the bond at the time of default, discounted using the risk-free interest rate.

For instance, if in the above example the default were to occur at the 6-month point rather than the 1-year point, then the LGD would be given as $108e^{-.05(.5)} - 40 = 105.33 - 40 = 64.33$.

Bootstrapping

Similar to what we did to determine interest rates for different maturities, we can calculate default probabilities for different time periods by using a sequence of bonds at different maturities to calculate the first year default probability, then the second year default probability, etc.

Risk-Free Rate

Note that the risk-free rate is used in both of the methods presented above for estimating default probabilities – either as the basis for calculating the yield spread or as the basis for discounting the promised cash flows. A common assumption is that the risk free rate is equal to the LIBOR/swap rate less 10 basis points.

Using Asset Swaps

Another approach to estimating the spread of bond yields over the LIBOR/swap rate is to use asset swap rates. Asset swaps exist so that investors can simply take on the credit risk portion of a corporate fixed rate coupon bond and not assume the inherent interest rate risk associated with the principal.

In an asset swap (assuming the bond trades at par), the swap buyer pays the promised fixed coupon rate on a specified bond to the swap seller in exchange for payments equal to LIBOR plus a spread[1]. The maturity of the swap is the same as the maturity of the underlying bond.

Earlier, we said we could solve for the default probability using the differences between the risk-free value of the bond, G and the actual value of the bond, B. But if an asset swap exists, we could also just use the risk-free present value of the fixed payments from/to the swap seller, since that should be equal to the expected present value of the default losses.

[1] If the bond trades below par, the difference is paid by the swap buyer to the swap seller at the outset. If the bond trades above par, the difference is paid by the swap seller to the swap buyer at the outset.

Note: The text's description of asset swaps is not very clear. See the GoldfarbSeminars.com download page for a more complete description of how an asset swap works and an explanation for why the present value of the spread is equal to the difference in present values of the risk-free bond and the corporate bond.

Comparing Default Probabilities

Estimates of default probabilities based on historical data are much lower than estimates based on bond prices.

One of the main reasons for this has to do with the distinction between *risk neutral* and *real world* default probabilities. The default probabilities that are derived based on bond prices are actually risk-neutral probabilities. They were derived assuming that the proper discount rate for the expected cash flows was the risk free rate. The historical probabilities are the real world default probabilities. These can be used only if the cash flows are discounted at a risk-adjusted rate.

Why do these differ so much? As the text shows, differences in the risk neutral and real world default probabilities translate into an expected excess return. There are several possible explanations for why traders would have an expected excess return:

- Liquidity Premium — Corporate bonds are relatively illiquid and so there may be a liquidity risk premium.
- Conservatism — Perhaps bond traders use subjective default probabilities that exceed the historical estimates, allowing for scenarios worse than what had been experienced over the time period used to measure the historical default probabilities.
- Systematic Risk and Credit Contagion — The fact that default rates fluctuate, perhaps systematically, gives rise to a source of risk that traders want to be compensated for.
- Skewness — Bond returns are skewed (limited upside, large downside) and therefore it is more difficult to diversify the risk in a bond portfolio. This could give rise to a risk premium for non-systematic risk.

So which rates should be used, the real world or the risk neutral ones? If the goal is to price a bond or value a credit derivative, it doesn't matter much which method we use (risk neutral probabilities and risk free rates or real world probabilities and risk adjusted rates). However, since risk-neutral methods are generally used to price derivatives, the risk neutral default rates should be used. But if we want to make probabilistic statements about the cash flows then we have to use real world probabilities.

Using Equity Prices to Measure Default Probabilities

One flaw in the two approaches discussed already is that they are both based on credit ratings, which suffer from a few problems, including the fact that they are not revised frequently.

Therefore, Hull outlines an approach whereby bond default probabilities can be inferred from the prices at any point in time of the issuing company's stock. The method is based on a Merton's characterization of equity as a call option on the company's assets, with a strike price equal to the face value of the debt.

Consider this analogy for a moment. When a firm is owned entirely by equity holders, they own all of the assets of the firm — the physical assets plus the income that those assets produce over the life of the company. If the equity holders issue debt (i.e. borrow money), then the equity holders no longer own all of the value of the firm, V. Instead, they own the excess of the value of the firm and the debt that they have to repay at time T, denoted D.

In other words, the value of the equity at time T is the residual of the firm value after the debt is repaid, if positive:

$$E_T = \max(V_T - D, 0)$$

which looks like a call on the value of V_T with a strike price of D.

It might help to think of this analogy as follows. When the equity holders borrowed the present value of D, they gave all of the assets of the firm to the bondholders, who will keep them if the debt is not repaid. However, by repaying the debt at time T, the equity holders have the right to buy back the assets of the firm by paying D. If $V_T < D$ on that date, they will not buy the assets back and will let the bondholders keep the assets. In other words, they will default.

Now, using this analogy, we can use the face value of the debt and the (known) market value of the equity to estimate the probability distribution of the value of the firm's assets at time T. From this, it is relatively simple to determine the probability that $V_T < D$, which represents the probability of default. We could also infer the probability of V_T being below any other value and therefore determine the probability that the debt is downgraded to any credit rating other than default too.

Mathematical Details of the Merton Model

Note: In what follows I will assume that students are familiar with the basic application of the Black-Scholes formula for pricing a call option and with Ito's Lemma. For those who are not familiar with these topics, please refer to the summary of the key option pricing results included in this manual.

Using the Black-Scholes formula, we can specify the value of the equity of the firm today in terms of the value of the assets and their volatility as follows:

$$E_0 = V_0 N(d_1) - De^{-rT} N(d_2)$$

$$d_1 = \frac{\ln(V_0/D) + (r + \frac{1}{2}\sigma_V^2)T}{\sigma_V \sqrt{T}}$$

$$d_2 = d_1 - \sigma_V \sqrt{T}$$

Notice some important subtleties. First, the underlying asset here is V, the value of the firm's assets not the value of its stock. Second, the volatility term σ_V is the volatility of the assets, not the volatility of the stock. Also, T represents the date the debt is due to be repaid, which in this case has to be assumed to be a zero coupon bond with one maturity date.

But the rest is trivial. The (risk neutral) probability of default is just the risk neutral probability that the option will NOT be exercised. The quantity $N(d_2)$ in the Black-Scholes formula can be interpreted as the (risk-neutral) probability that the call option WILL be exercised, which makes the probability that it will not be exercised:

$$\text{Probability of Default} = 1 - N(d_2) = N(-d_2)$$

And now for the final detail. It is obvious where to get most of the inputs:

- E_0 is just the current value of the equity which we can observe based on the current stock price,
- D is just the amount of debt to be repaid at time T,
- r is the risk free rate.

But where do we get the current value of the asset, V_0, and the volatility of the assets, σ_V? We so far have only one equation and two unknowns.

To get another equation, we can use Ito's Lemma to see that:

$$\sigma_E E_0 = N(d_1)\sigma_V V_0$$

Derivation

For those interested, it is relatively trivial to apply Ito's Lemma to see the above relationship between the volatility of the equity (σ_E) and the volatility of the assets (σ_V).

To derive this relationship, note that according to the Merton model, the equity (E) is a call option on the value of the firm, V and can be written as:

$$E = G(V)$$

where

$$G(V) = \max(V - K, 0)$$

If we assume that V follows geometric Brownian motion, Ito's Lemma tells us that E will also follow geometric Brownian motion and can be written as:

$$dE = \mu E\,dt + \sigma_E E\,dz$$

We also know from Ito's Lemma that this can be written in terms of the process followed by V, as:

$$dG = \left(\frac{\partial G}{\partial V}\mu V + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial V^2}\sigma_V^2 V^2\right)dt + \frac{\partial G}{\partial V}\sigma_V V\,dz$$

We aren't interested in the drift term, but we can see by equating the volatility terms that:

$$\sigma_E E = \frac{\partial G}{\partial V}\sigma_V V = N(d_1)\sigma_V V$$

This gives us two equations and two unknowns. We can solve for V_0 and σ_V and then we can determine the probability distribution for V_t.

Option Delta

In the derivation above I quickly glossed over the that that the expression for the partial derivative of the option value with respect to the value of the assets is equal to $N(d_1)$, or

$$\frac{\partial G}{\partial V} = N(d_1)$$

To derive this, we just need to use calculus to take the partial derivative of the value of the call option on the assets. If you want to see that derivation, refer to the file on the downloads page of my website (Derivation of the Option Greeks.pdf).

Practice Questions

Question 1. Suppose a one-year zero coupon corporate bond with credit risk has an annually compounded yield of 5.8% and the annually compounded risk free yield is 5.1%. If the recovery rate is 40%, what is the approximate annual hazard rate?

Solution. The formula given for the approximate annual hazard rate is:

$$\bar{\lambda} = \frac{s}{(1 - R)}$$

Here, the spread was given as 5.8% – 5.1% = .7% and the recovery rate was 40%. This translates into a hazard rate of:

$$\bar{\lambda} = \frac{.7\%}{(1 - .40)} = 1.167\%$$

Question 2. Assume that the continuously compounded risk free rate (based on the LIBOR/swap rate) for a three-year zero-coupon bond is 5.0% and that the continuously compounded yield on a 3-year zero-coupon corporate bond rated AA is 5.7%. What is the expected loss from default for the corporate bond? Assume both bonds have a face value of $100.

Solution. First, the price of the risk-free bond is:

$$G = \$100e^{-.05(3)} = \$86.0708$$

The price for the corporate bond is:

$$B = \$100e^{-.057(3)} = 84.282$$

This means that investors are willing to pay $1.7886 more for a risk free bond than a corporate bond with credit risk. This amount represents the present value of the expected loss from default.

Question 3. Assume that for the 3-year bond in the previous question, in the event of default the recovery rate is 40% of the face value of the bond and that default can only occur at maturity. Determine the unconditional 3-year default probability, $Q(3)$, at time zero.

Note that we are looking for the unconditional default probability over a three year period since a zero coupon bond cannot default prior to its maturity. This is therefore a cumulative default probability through year 3.

Solution. If the bond defaults at maturity, the risk free value will be $100 and the recovery amount will be $40. This gives us a loss given default of $60. The risk free present value of this is:

$$\$60e^{-.05(3)} = \$51.64$$

The present value of the expected default loss is therefore equal to $\$51.64Q(3)$. From the previous question, we saw that the expected loss from default was $1.7886.

Setting these two amounts equal to each other, we can solve for $Q(3)$:

$$Q(3) = \$1.7886/\$51.64 = 3.463\%$$

Another way to solve this same problem, when the situation is simplified using zero coupon bonds, is to simply solve for the unconditional default probability, $Q(3)$, which sets the bond price equal to the expected (risk-free) present value of the payments.

Recall that if the bond does not default it pays $100; if it defaults with probability $Q(3)$ it pays 40%($100) = $40. Therefore, we simply solve for $Q(3)$:

$$100e^{-.057(3)} = [(1 - Q(3))(100) + Q(3)(40\%)(100)]e^{-.05(3)}$$

Solve for $Q(3) = 3.4635\%$.

Question 4. What is the average instantaneous default intensity (average instantaneous hazard rate) for the bond in the previous question?

Solution. Notice that the default probability in the previous question reflects the cumulative probability of default by year 3. In the text, this was denoted $Q(3) = 3.4635\%$. Using the following formula:

$$Q(t) = 1 - e^{-\bar{\lambda}t}$$

we can solve for the average instantaneous default intensity, $\bar{\lambda}$, as follows:

$$\begin{aligned}\bar{\lambda} &= -\frac{1}{3}\ln[1 - Q(3)] \\ &= -\frac{1}{3}\ln[1 - 3.463\%] \\ &= 1.17\%\end{aligned}$$

Question 5. Suppose you had a three-year semi-annual coupon bond with \$100 face value paying a coupon of 7% and trading at a continuously compounded yield of 6.8%. Assume the continuously compounded risk free yield is 5% and in the event of default the recovery is 30% of the face value. Also assume that default can only occur at the end of each year (just before the coupon is paid) and that the unconditional annual probability of default, Q, is the same each year. Determine the value of Q.

Solution. Notice that I made the default only possible at the end of each year just to eliminate some of the calculations. More importantly, I also set up the question so that the annual probability of default is constant for each of the three years.

The first step is to determine the risk-free price of the bond and the actual price of the bond. The difference between these two amounts will represent the present value of the expected default losses.

The bond pays \$3.5 in interest every six months and \$103.5 at maturity, and therefore its risk free price is:

$$\begin{aligned}G &= 3.5e^{-.05(.5)} + 3.5e^{-.05(1)} + 3.5e^{-.05(1.5)} \\ &\qquad + 3.5e^{-.05(2)} + 3.5e^{-.05(2.5)} + 103.5e^{-.05(3)} \\ &= 105.33\end{aligned}$$

Then, using the 6.8% bond yield given the market price of the bond is:

$$\begin{aligned}B &= 3.5e^{-.068(.5)} + 3.5e^{-.068(1)} + 3.5e^{-.068(1.5)} \\ &\qquad + 3.5e^{-.068(2)} + 3.5e^{-.068(2.5)} + 103.5e^{-.068(3)} \\ &= 100.22\end{aligned}$$

This tells us that the total expected present value loss from default is:

$$G - B = 105.33 - 100.22 = \$5.11$$

Since default can only occur at the 1, 2 or 3 year points, we need to know the amount of *loss* at each of those points. The loss is defined as the risk free value of the bond at that point less the recovery in the event of default.

Let's start with the risk free values of the bond at year 1, 2 and 3. If we are sitting at the end of year 1 right before the coupon is paid, then we have cash flows of \$3.5 right now, then \$3.5 for 3 more periods and \$103.5 at maturity. The value of this using the risk free yield of 5.0% is:

$$\begin{aligned} F_1 &= 3.5 + 3.5e^{-.05(.5)} + 3.5e^{-.05(1)} + 3.5e^{-.05(1.5)} + 103.5e^{-.05(2)} \\ &= 107.141 \end{aligned}$$

Similar calculations are done for the other two possible default times.

Then, the following table can be filled in. The loss given default (LGD) reflects the risk free value at each point less the recovery of 30%(\$100) = \$30. Multiplying by the default probability, assumed to be a constant Q in each period, and the risk free discount factor, we get a total present value of expected loss equal to $204.38Q$.

TABLE 3. Calculation of Expected Loss from Default

T	Default Prob	Recovery	Risk-Free Value	LGD	Discount Factor	PV of Exp Loss
1	Q	30.000	107.141	77.141	0.9512	73.379Q
2	Q	30.000	105.366	75.366	0.9048	68.194Q
3	Q	30.000	103.500	73.500	0.8607	63.262Q
						204.835Q

To find Q we simply set $204.835Q = 5.11$ and solve for $Q = 2.495\%$.

Question 6. Suppose that the bond in the previous question had a current rating that showed historical default probabilities of only 0.50%. What could explain the discrepancy between the 0.50% historical default probability and the 2.49% default probability calculated above?

Solution. The main point to note is that the 2.49% estimate is a *risk-neutral* probability and these generally are much higher than the historical probabilities. Among the explanations for this are:

- Liquidity Premium — Corporate bonds are relatively illiquid and so there may be a liquidity risk premium.
- Conservatism — Perhaps bond traders use subjective default probabilities that exceed the historical estimates, allowing for scenarios worse than what had been experienced over the time period used to measure the historical default probabilities.

- Systematic Risk and Contagion — The fact that default rates fluctuate, perhaps systematically, gives rise to a source of risk that traders want to be compensated for.

- Skewness — Bond returns are skewed (limited upside, large downside) and therefore it is more difficult to diversify the risk in a bond portfolio.

Question 7. Assume a Baa-rated bond has one-year unconditional default probability of .177% and a two-year unconditional default probability of .495%. Calculate the average hazard rate during year one and the average hazard rate during year two for this bond.

Solution. Notice that this question did not ask for the average hazard rate over the entire two-year period. Instead, it asked for the hazard rate during the second year.

To find the hazard rate during the second period, we use the fact that if λ_i is the hazard rate for period i then $e^{-\lambda_i \Delta t}$ is the probability of *surviving* during the time period Δt. Therefore,

$$1 - .495\% = (1 - .117\%)e^{-\lambda_2(2-1)}$$

That is to say, the two-year survival probability is equal to the product of the one-year survival probability and the probability of surviving during the second period. We can solve this to get $\lambda_2 = .379\%$.

Question 8. A company has a one-year bond outstanding with a \$100 face value and 8% coupons paid semi-annually. You have used the market value of the bond to estimate the average hazard rate during the year and found it to be $\lambda_1 = 3.00\%$ by setting the expected present value of default losses equal to the difference between risk-free value of the bond and the market value of the bond. Determine the market value of the bond under the assumptions that default can only occur on the coupon payment dates, just prior to the coupon being paid, and that in the event of default the recovery is equal to 40% of the face value of the bond. Assume the risk-free yield is 5%, continuously compounded for all maturities.

Solution. The first step is recognize the equation that needs to be solved. We know that the present value of expected default losses is found by summing, across all possible default dates, the product of the probability of default and the present value of the loss given default. To solve for the value of the bond, we are going to subtract the present value of the expected default losses from the risk-free value of the bond.

Start with the default probabilities. We are given the hazard rate during year 1, so that means we can use:

$$1 - Q(t) = e^{-\bar{\lambda}t}$$

to find $Q(.5) = 1 - e^{-.03(.5)} = 1.489\%$. That is the probability of default in the first 6-month period. Then, following the logic of the previous question, we can find the probability of default through the second 6-month period as $Q(1) = 1 - [(1 - 1.489\%)e^{-.03(1-.5)}] = 2.955\%$. Using that, we can find the probability of default during the second 6-month period as $2.955\% - 1.489\% = 1.467\%$.

Note that I could have just calculated $Q(1) = 1 - e^{-.3(1)}$. The reason I did not do that is because if the hazard rates change from period to period, you need to be more careful and do it the first way I did it.

We now have the two probabilities we need. Now we can proceed to calculating the loss given default at each possible default point. If default occurs just before the first coupon, the risk-free value of the bond at that point is $G_{.5} = 4 + 104e^{-.05(.5)} = 105.43$. If default occurs just before the second coupon, the risk-free value of the bond at that point is $G_1 = 104$.

Then, we just have to subtract the recovery rate, multiply by the risk-free present value factor and multiply by the default probability to get the present value of expected default losses of 1.84, as shown below:

Default Time	No-Default Value	Recovery if Default	Loss Given Default	PV of Loss	Prob of Survival	Prob of Default	Expected Loss
0.50	105.43	40.00	65.43	63.82	0.9851	0.01489	0.95
1.00	104.00	40.00	64.00	60.88	0.9704	0.01467	0.89
							1.84

Finally, the question asked about the market price of the bond. If it were risk-free, the bond would be worth 102.83. Since the present value of the expected default losses is 1.84, the value of the bond is 102.83 − 1.84 = 100.99.

Question 9. Using the same information as in the previous question, assume the hazard rate during the second year is 3.5% and that the same company has issued a 2-year bond also with \$100 face value and 8% coupons paid semi-annually. The risk-free rate is 5% continuously compounded for all maturities and the recovery rate is 40% of the face value of the bond. Determine the value of the two-year bond.

Solution. Using the same approach as the previous question:

Default Time	No-Default Value	Recovery if Default	Loss Given Default	PV of Loss	Prob of Survival	Prob of Default	Expected Loss
0.50	108.19	40.00	68.19	66.51	0.9851	0.0149	0.99
1.00	106.83	40.00	66.83	63.57	0.9704	0.0147	0.93
1.50	105.43	40.00	65.43	60.70	0.9536	0.0168	1.02
2.00	104.00	40.00	64.00	57.91	0.9371	0.0165	0.96
							3.90

The risk-free value of this 2-year bond is 105.52. Based on that, the value of the bond is 101.62.

You should have noticed that this question is identical to the previous one. The only twist is that you need to be careful when calculating the default probabilities because the hazard rates vary from the first year to the second year.

Question 10. In the previous two questions, we assumed that default could only occur right before the coupon is paid. Describe how the analysis would change if I had said that default can occur at any point during the 6-month period in between coupon dates.

Solution. The default probability that we calculated for each period really represents the probability of default at any time during the period. If defaults can occur at any time, then an easy way to modify the calculation is to just calculate the loss given default at the mid-point of the period. So, rather than calculate the forward risk-free price of the bond as of 6 months, 12 months, etc. we calculate it at 3 months, 9 months, etc. When we discount the loss given default, we also use 3 month, 9 month, etc. discount factors. We do not change the default probabilities or any other aspects of the calculation. We merely use a representative date during the period to calculate the loss given default.

Here is how the previous question would look with this change:

Default Time	**No-Default Value**	**Recovery if Default**	**Loss Given Default**	**PV of Loss**	**Prob of Survival**	**Prob of Default**	**Expected Loss**
0.25	108.19	40.00	68.19	67.34	0.9851	0.0149	1.00
0.75	106.83	40.00	66.83	64.37	0.9704	0.0147	0.94
1.25	105.43	40.00	65.43	61.47	0.9536	0.0168	1.03
1.75	104.00	40.00	64.00	58.64	0.9371	0.0165	0.97
							3.95

Question 11. Suppose a company has \$10 million in equity value (based on their current market price and the number of shares outstanding) with an equity volatility of 60%. It also has \$10 million in zero-coupon debt outstanding with 2 years to maturity. The risk free rate is 5%. You want to be able to estimate the current value of their assets and the volatility of their assets. Show the two equations you would solve simultaneously to estimate these two values? *Do not actually solve them, just show the equations.*

Solution. The value of a firm's equity can be viewed as a call option on the firm's assets with a strike price equal to the debt to be repaid. Using the Black-Scholes formula with the appropriate changes in notation:

$$E_0 = V_0 N(d_1) - De^{-rT} N(d_2)$$

$$d_1 = \frac{\ln(V_0/D) + (r + \frac{1}{2}\sigma_V^2)T}{\sigma_V \sqrt{T}}$$

$$d_2 = d_1 - \sigma_V \sqrt{T}$$

However, we have one equation with two unknowns and so we need another equation in terms of the two unknowns. The other equation we can use relates the stock price volatility to the asset value volatility, as follows:

$$\sigma_E E_0 = N(d_1)\sigma_V V_0$$

Note that the question did not ask you to solve for V_0 and σ_V because it isn't practical to do this by hand on the exam. You would have to use something like Excel's Solver or do a time-consuming and very difficult two-dimensional Newton-Raphson numerical procedure.

Nonetheless, in this case,

$$E_0 = \$10 \text{ million}, D = \$10 \text{ million}, r = 5\%, T = 2, \sigma_E = 60\%$$

Using these values, the current value of the assets and the volatility of the assets can be found to equal:

$$V_0 = 18.905 \text{ million}$$
$$\sigma_V = 32.87\%$$

Question 12. The previous question indicated that the current value of the assets was \$18.905 million and the volatility was 32.87% for the company in the previous question. Determine the current market value of the debt and the probability of default according to the Merton Model.

Solution. First, if the market value of the assets is \$18.905 million and the market value of the equity is \$10 million, then the market value of the debt must be $\$18.905 - \$10 = \$8.905$ million.

Using the values from above, $d_2 = 1.3527$.

As discussed in a Hull chapter that is not on the syllabus, the risk neutral probability of exercising the option is $N(d_2) = .9119$. That means that the probability of NOT exercising the option, which is to say that the probability of the equity holders not repaying the debt to buy back the assets of the firm, is:

$$\text{Probability of Default} = 1 - .9119 = 8.81\%$$

Question 13. Using the information from the previous question, what is the expected recovery rate on the debt?

Solution. We saw that the market value of the debt is \$8.905 million. If the debt were risk free, it would be worth:

$$G = 10,000,000e^{-.05(2)} = \$9.048 \text{ million}$$

The expected present value default loss is therefore:

$$\$9.048 - \$8.905 = \$143,000$$

Another way to calculate the expected present value loss from default is to multiply the loss given default (LGD) by the default probability and then discount that at the risk free rate.

If P is the principal, Q is the default probability and R is the recovery rate:

$$\begin{aligned}\text{Expected PV Loss from Default} &= (P - [(1-Q)P + QRP])\,e^{-rt} \\ &= QP(1-R)e^{-rt}\end{aligned}$$

The quantity $P(1-R)$ above is the loss given default.

Therefore,

$$\$143,000 = Q(\$10 \text{ million})(1-R)e^{-.05(2)}$$

Plug in 8.81% for Q from the previous question and solve for $R = 82\%$.

Question 14. Suppose the LIBOR/swap curve is flat at 6% with continuous compounding. A five-year bond with \$100 face value and 5% coupons paid semi-annually has a price of \$90.00. Determine the *asset swap spread* assuming one party (the swap buyer) paid \$10 (the difference from par) at inception plus the promised coupons each period over the life of the bond and the other party (the swap seller) paid LIBOR plus the asset swap spread.

Solution. If the bond had no credit risk, it would have been worth the present value of the cash flows discounted at the 6% LIBOR/swap rate, $G = \$95.358$. The calculations are shown in the table below:

TABLE 4. Risk-Free Value of the Bond

T	Coupon	Face Value	PV Factor	PV(Cash Flow)
0.50	2.50	0.00	0.970	2.426
1.00	2.50	0.00	0.942	2.354
1.50	2.50	0.00	0.914	2.285
2.00	2.50	0.00	0.887	2.217
2.50	2.50	0.00	0.861	2.152
3.00	2.50	0.00	0.835	2.088
3.50	2.50	0.00	0.811	2.026
4.00	2.50	0.00	0.787	1.967
4.50	2.50	0.00	0.763	1.908
5.00	2.50	100.00	0.741	75.934
			8.510	95.358

The bond was said to be worth $B = \$90.00$. This tells us that the expected present value of the default costs must be:

$$G - B = \$95.358 - \$90.00 = \$5.358$$

Since the swap buyer will suffer the default costs but will also receive a net spread equal to s each period for five years, the present value of the spread must equal \$5.358. The sum of the present value factors in the table above gives us the value of a \$1 annuity paid every six months for five years, so that means that $8.510s = 5.358$.

We can solve for $s = \$0.63 = .63\%$ of the par amount paid every six months. That is an asset swap spread of 126 basis points per annum.

The textbook left out some important aspects of the asset swap that makes it hard to see why the expected default costs and the value of the spread payments are supposed to be equal to each other. Some additional details are in a file on the GoldfarbSeminars.com download page if you are interested.

Recommended Textbook Problems

I strongly recommend working all of the end-of-chapter questions from the Hull textbook. But due to time constraints this may not be feasible. Therefore, at a minimum you should review all of the numerical examples that appear throughout the main text and the following end-of-chapter questions (especially the ones in bold):

1, 3, 5, **12, 13**

Hull Chapter 23: Credit Risk in Derivatives Transactions

Adjusting Derivative Prices for Default Risk

Over the counter derivative transactions involve potential credit risk of the counterparties that may need to be taken into account when valuing a derivative position. This is more complicated than it may first appear, since the loss in the event of default depends upon which party to the derivative owes under the derivative when the default occurs.

Contracts That Are Always Liabilities

If the derivative is always a potential liability for you, such as a short position in an option, then you have no credit risk from your counterparty (they have credit risk from you). In that case, there is no need to make any adjustment.

Contracts That Are Always Assets

Some derivatives are always an asset to you, such as a long position in an option. That is to say, your counterparty is always the party who owes money under the contract. In this case, the credit risk adjustment is the same as it is for a bond with credit risk.

If q_i is the probability of default at time period i and R is the recovery rate, we merely adjust the risk-free value today, f_0, by the expected present value of the losses from default. The expected value of the losses from default are simply the risk free value (f_0) times the cumulative product at each point in time at which default can occur of the default probability times one minus the recovery rate:

$$\text{Expected Losses from Default} = f_0 \sum_{i=1}^{n} q_i(1-R)$$

The adjusted value of the derivative, f_0^*, will simply reflect the unadjusted value, f_0, less the expected losses from default:

$$f_0^* = f_0 - f_0 \sum_{i=1}^{n} q_i(1-R) = f_0 \left[1 - \sum_{i=1}^{n} q_i(1-R) \right]$$

Notice that the relationship between the adjusted derivative price and the unadjusted derivative price will be the same as the relationship between a risky bond (with the same maturity) issued by the same counterparty and a risk-free bond. That is, if $B_0 = e^{-yT}$ is the risk-free value of a bond with a payoff of \$1 at time T and $B_0^* = e^{-y^*T}$ is the value of a bond issued by the counterparty with a payoff of \$1 at time T, then:

$$f_0^* = f_0 \frac{B_0^*}{B_0}$$

Contracts That Can Be Assets or Liabilities

For a contract that can be either an asset or a liability the adjustment is trickier. In this case, you need to determine the value, v_i, of the claim amount on the derivative position in the event of default. This claim amount is just the (risk-free) value of the derivative position at that point in time, if it is positive, or zero.

$$v_i = e^{-rt_i}\hat{E}[\max(f_i, 0)]$$

where $\hat{E}[\cdot]$ represents the risk neutral expected value.

The resulting formula for the expected losses from default is essentially the same as above in the case where the derivative is always an asset, except that instead of being able to factor out the current value of the derivative (f_0) we have to, at each possible default point, reflect only the value of the derivative if it is positive:

$$\text{Expected Losses from Default} = \sum_{i=1}^{n} q_i(1-R)v_i$$

This expected loss from default on a derivative instrument is referred to as the *credit value adjustment (CVA).*

Example

Note that it is unclear from the Learning Objectives to what extent you are responsible for the discussion regarding adjusting derivative prices for default risk. The numerical problems that would be associated with this topic could require you to rely on derivative pricing formulas that are not on the syllabus, such as the value of a put option on a foreign currency. You may want to be prepared for this, but the learning objectives only reference being able to "describe the credit risk in derivatives transactions and various mechanisms to manage the risk." The example here and in the practice problems may be more than you need to know so feel free to skip this.

As an example of this, consider (from Question 23.15 in the text) a one year forward contract to sell $100 for AUD 150 (150 Australian dollars). This is equivalent to 150 forward contracts on the Australian dollar at a forward exchange rate of 1.50 AUD per U.S. dollar, or equivalently AUD 0.6667 per U.S. dollar. The one year U.S. dollar risk-free rate is 5% and the counterparty can borrow at a one-year U.S. dollar interest rate of 6%. If the forward exchange rate is currently 1.50 AUD per U.S. dollar and the exchange rate volatility is 12% per annum, what is the present value of the expected default costs if default is only taken into account at the maturity of the forward contract?

To begin, we need the cumulative probability of default for the counterparty in one year and the recovery rate if they do default. In reality we only need the product $q_i(1-R)$, so there's a simple way to use their borrowing rate to get this.

We know that a risk free bond that pays $100 in one year is worth:

$$G = \$100e^{-.05(1)} = \$95.123$$

We also know that if the counterparty issued a bond that paid \$100 in one year, people would pay:

$$B = \$100e^{-.06(1)} = \$94.176$$

That means that the present value cost of default is:

$$G - B = \$95.123 - \$94.176 = \$0.947$$

When this is divided by the risk free value of the bond, we get the expected loss from default:

$$q_i(1 - R) = \frac{0.947}{95.123} = .009950$$

Now all we need is the present value of the claim amount. In this case, we only care about the case where the other party owes us money. That will occur when the value of the AUD150 they are going to give us is more than the \$100 we are going to give them at the maturity. In other words, that occurs when AUD 150 times the exchange rate in dollars (S_T) is greater than \$100. The quantity v is then the value of the payoff:

$$v = \max(150S_T - 100, 0) = 150[\max(S_T - 0.6667, 0)]$$

Notice that to value this quantity we simply need to value 150 call options on the AUD exchange rate with a strike price of \$0.6667. This is because the quantity $\max(S_T - 0.6667, 0)$ is simply a call option on the AUD with a strike price of \$0.6667. I won't show the steps here, but this is equal to \$4.545 (see Hull Chapter 15 notes for the formulas for a call option on currencies).

Finally, the expected loss from default is simply:

$$v_i q_i(1 - R) = (4.545)(.009950) = .04522$$

Additional Considerations

Note that this discussion somewhat simplified matters because the claim in the event of default might also depend upon what other positions exist with the same counterparty, how netting is done, what collateral existed, whether the default probability and the claim amount are related, etc.

Credit Risk Mitigation

Three clauses in derivative contracts that are often used to mitigate exposure to credit risk include the following:

- Netting — The credit losses that result from the default of a particular counterparty can be mitigated through *netting*. The specific contractual provision stipulates that if the counterparty defaults on one transaction (presumably one in which they owe money to their counterparty) then they must default on all transactions with that counterparty, including those for which they have a positive net position. This prevents them from selectively defaulting on transactions where they owe money but keeping alive other derivative transactions in which they are owed money.

- Collateral Requirements — These are requirements for a counterparty to post collateral in the form of cash or marketable securities in the amount of their potential losses from one or more transactions. The magnitude of these requirements typically vary based on the current credit rating of the counterparty and may reflect simple percentages of the maximum possible exposure or more complicated formulas such as a Value at Risk measure.

- Downgrade Triggers — These are conditions included in contracts that force an early termination of the transaction if the counterparty credit rating drops below a threshold.

Practice Questions

Question 1. Suppose Party A and Party B entered into swap with the following terms. It is a 5-year swap with annual payments, Party A pays 10% annually compounded on notional of 50 million pounds sterling, Party B pays 5% annually compounded on $100 million, the initial exchange rate is 2.000, the volatility of the exchange rate is 15%, the sterling yield curve is flat 10% annually compounded and the dollar yield curve is flat at 5% per annum. Assuming that default can only occur at time $t = 2$ years, calculate the present value of the claim amount in the event of default of Party B.

Solution. Note that a currency swap is like one party holding a long position in one bond and a short position in another bond. The value of the swap then is just the difference in the values of these two bonds. At inception, since the yield curve is flat and equal to the coupon rates, both bonds trade at par.

Since we are interested in the value of these bonds at $t = 2$, we can use the method discussed in Hull Chapter 7 and assume that the forward rates are realized and value each side of the swap at the current risk free rates. Because of the flat yield curves, the forward rates will all equal the current rates and therefore on any coupon date, just before the coupon is paid, the bonds will always be worth par plus the accrued interest.

Therefore, if there is a default at $t = 2$, the forward value of the sterling bond will be 55 million sterling and the dollar bond will be $105 million.

Since Party A is receiving dollars and paying sterling, he is long the dollar bond and short the sterling bond. Therefore, his position in millions of USD is worth $105 - 55S_2$ where S_2 is the (unknown) exchange rate expressed in dollars per pound sterling at time $t = 2$.

The value of the claim in the event of default then is:

$$\max(0, 105 - 55S_2) = (55)\max(0, 105/55 - S_2)$$

The term inside the brackets is just the value of a sterling put option with a strike price of $105/55 = 1.91$.

The tricky part of this though is the maturity. Here, we are interested in the value of the put at time $t = 2$. Don't get confused by the 5-year maturity of the swap.

To value that put option, we need to determine the forward exchange rate at time $t = 2$. We first have to convert the annually compounded U.S. and Sterling interest rates to continuously compounded rates, $r = .0488$ and $r_f = .0953$. Then,

$$F = S_0 e^{(.0488-.0953)2} = 1.8223$$

Then, using the formulas from Hull Chapter 16 for the put option on a currency:

$$p = e^{-rT}[KN(-d_2) - FN(-d_1)]$$

$$d_1 = \frac{\ln(F/K) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

So, $d_1 = -.1132, d_2 = -.3254$ and $p = .1857$.

Now note that the put value above is just the portion of v_i equal to $\max(0, 105/55 - S_2)$. Therefore,

$$v_i = 55(.1857) = 10.214$$

Background: Mortgage Backed Securities and Related Securitizations

Mortgage backed securities (MBS) and related securitizations are discussed in a variety of chapters of BKM as well as other assigned readings on the syllabus. They play an important role in the Asset-Liability Management readings, serve as important background for understanding the mechanics of catastrophe bonds and played a critical role in the Financial Crisis of 2007–2009.

This reading will provide an introduction to this topic and draw from several of the assigned readings. Additional details that are specific to other readings will be discussed in the notes to those readings and will assume familiarity with this summary.

Mortgage Pass Throughs

These are the simplest types of MBS to understand. I will go into greater depth in describing these than may seem necessary because this will make it easier to understand the more complex variations discussed later.

Basics of Home Mortgages

When someone borrows money to purchase a house, and uses that house as collateral for the loan, the person is said to have a *mortgage*. The lender owns the mortgage and receives cash flows just as he would if he owned a traditional bond. So when thinking about the position of the lender, it is sometimes helpful to refer to this as a mortgage bond.

Historically most mortgages required large down payments of 10%-20%, carried a fixed rate of interest, matured in 30 years and had level payments on a monthly basis. In more recent years banks introduced adjustable rate mortgages and other innovations that introduced significant variation in the cash flow profiles of various mortgages. For the moment, we'll focus on conventional 30-year fixed rate mortgages with low loan-to-value ratios (where the amount borrowed is low relative to the value of the collateral).

An important characteristic of mortgages is the fact that the borrower is always free to prepay the mortgage early, for instance as would be necessary if they decided to move and no longer could pledge the house as collateral for their loan.

Risk to the Bank Issuing Mortgages

For the bank issuing that loan, there are two sources of risk that they face when providing a mortgage.

1. **Default Risk** — One risk is that the person doesn't repay the loan on time or in full. This default risk can be managed in a variety of ways, including the following:

a. Do not to allow the loan to exceed the value of the property or some fraction of the value of the property. Failure to pay allows the bank to foreclose on the house and sell it for its market value in order to repay the outstanding balance.

b. Obtain mortgage insurance on the borrower that would transfer this default risk to a third party.

c. Carefully underwriting the borrower's ability and *willingness* to repay, etc.

2. **Prepayment Risk** — The other source of risk is that the homeowner prepays faster or slower than expected. The bank may not want to get the principal back early if, for instance, interest rates had fallen. But this is precisely when homeowners are likely to repay early, so that they can take advantage of better rates available (refinancing) or by deciding that they can afford a bigger home now that the rates are lower and their monthly payments will be reduced.

   Of course, they may also prepay simply because they want to move for reasons that have nothing to do with changes in rates, such as job-related relocation, desire for bigger/smaller house, etc. These effects are of less concern here because they are not highly correlated with changes in interest rates and so they may occur when rates have moved in such a way that prepayment is good for the bank.

Managing the Risk

Banks issuing mortgages have a variety of ways to manage these credit and prepayment risks. Of course they could simply sell the mortgage to another party. Or, they could stop short of actually selling the mortgage (e.g. the bank might want to maintain the customer relationship) and simply enter into an agreement with an investor to "pass through" all of the cash flows received by the borrower. This would effectively transfer all of the risk and cash flows as in a true sale.

But of course, this other party would have the same concerns about prepayment — there will still be a significant amount of uncertainty regarding when the principal for a single mortgage will be paid.

But by *pooling* several mortgages together, much of this uncertainty can be diversified away, making the overall cash flows more predictable. If one of the loans defaults or prepays faster than expected, the actual cash flows will not vary as greatly from the expected cash flows as if this were a single mortgage. The collective payments from each mortgage are then passed through to various investors in the form of bonds which receive portions of the cash flows from the underlying mortgages. In addition to diversifying default and prepayment risk, these pass through securities allow investors to own smaller and more liquid portions of mortgages.

Important Note About Credit Risk in Pass Through Securities

For the time being, we will assume that the default risk is not material and assume that the only risk to investors in the pass through is the prepayment risk. This is somewhat reasonable because, in addition to the steps noted above that the bank can take to minimize credit risk, historically quasi-government agencies such as Fannie Mae and Freddie Mac (which are now directly owned by the US federal government) would purchase mortgages that conformed to certain underwriting guidelines directly from originating banks and then issue the pass through security to investors with the principal and interest payments guaranteed against default risk by Fannie Mae or Freddie Mac. So neither the issuing bank, which has already sold the mortgage to Fannie Mae or Freddie Mac, nor the investors who received the quasi-government guarantee against default risk bore the default risk in mortgage pass through securities.

However, beginning in 2007 there were major problems in the MBS market that resulted from the proliferation of loans to homeowners with substantial credit risk (so-called sub-prime and Alt-A borrowers). These loans typically did not conform to the Fannie Mae and Freddie Mac standards and thus were not guaranteed. In these instances, the credit risk was also passed through to investors. We will return to this very important issue later in this discussion when we address asset-backed securities (ABS) and sub-prime residential mortgage backed securities.

Numerical Example — 10% Pass Through of a Single Mortgage

Assume a mortgage has scheduled monthly payments of interest and principal over the 360 month life of the mortgage as shown in Table 1.

TABLE 1. Sample Mortgage Cash Flows

Month	Beg Balance	Interest	Scheduled Principal
1	100,000.00	666.67	67.10
2	99,932.90	666.22	67.55
3	99,865.36	665.77	68.00
⋮	⋮	⋮	⋮
360	728.91	4.86	728.91

Now consider the cash flows in month 3 to an investor who owns 10%, or \$10,000 worth, of this mortgage. He expects to receive 10% of the total cash flows or 10%(665.77 + 68.00) = 73.38.

If the borrower unexpectedly prepays the full balance of the mortgage at the end of month 3, then their total payment will be the interest they owed, 665.77, plus the full principal amount outstanding of \$99,865.36 (note this is the beginning principal — you can also think of it as they paid 68.00 of the principal in the month and then made an extra principal payment of \$99,797.36 so that their month 4 beginning principal is zero).

The investor gets 10% of this total payment, or \$100,531(10%) = \$10,053.10. This is 137 times more than what was scheduled.

Numerical Example — 1% Pass Through of a Mortgage Pool

Instead of a single mortgage, assume that we pooled together 10 identical mortgages and created pass through securities from these. An investor could still own \$10,000 worth of securities by owning just 1% of this total pool. The scheduled cash flows would be the same as before — the pool is 10 times larger but we own only one-tenth the percentage that we owned before.

Now assume that only 1 of the 10 mortgages prepays the full balance at the end of month 3. What are the scheduled and actual cash flows in month 3?

Since the pool cash flows in month 3 are scheduled to be 10(733.77) = \$7,337.70 and the investor owns 1% of the pool, their scheduled cash flows are the same as before, \$73.38. Now if the borrower prepays unexpectedly, then the cash flows for the entire pool will be 9(733.77) + \$100,531 = \$107,134.93. The investor will get 1% of this amount, or \$1,071.35. This is 14.6 times what was scheduled, which is considerably closer to the scheduled cash flow (i.e. far less volatile) than when the investor owned the same dollar amount of a single mortgage. Obviously if there were thousands of mortgages in the pool, the prepayment risk would be even more diversified.

Notice that this is a practical example of the discussion of diversification, risk pooling and risk sharing in BKM Chapter 7.

Collateralized Mortgage Obligations (CMO)

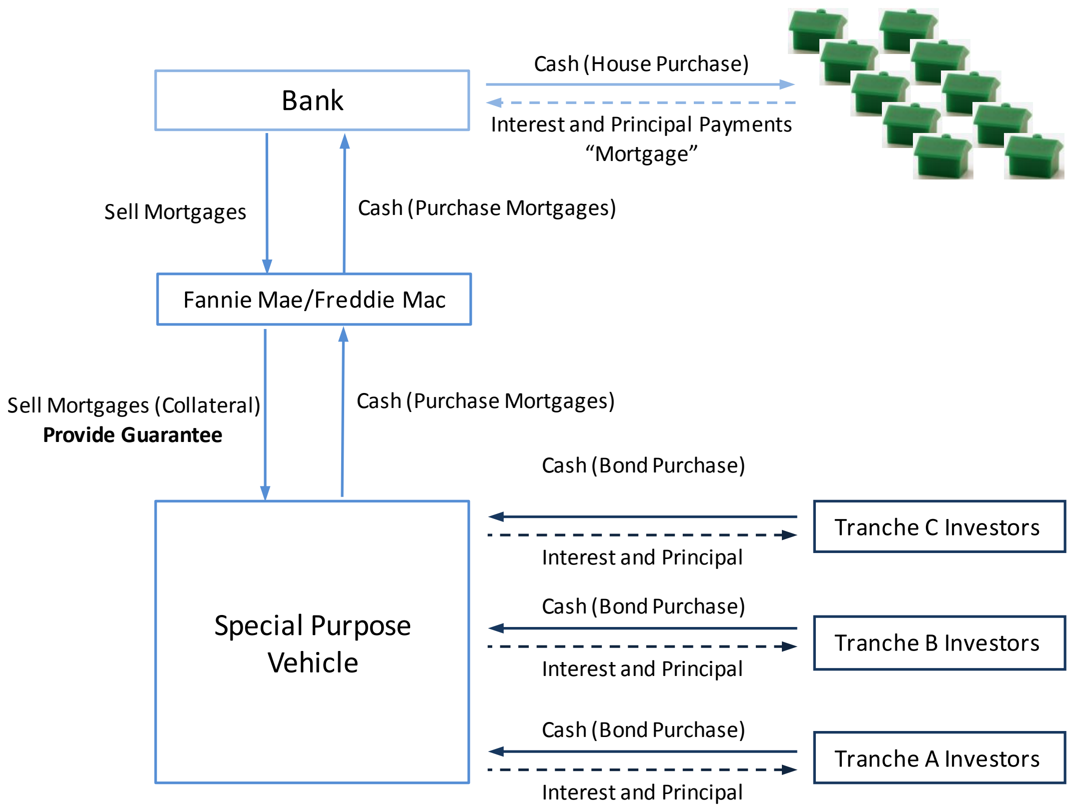
A CMO is similar to a pass through, but the interest and principal payments are not passed through to all investors on a pro-rata basis. Instead, various classes or tranches are created, whereby prepayments are allocated according to a prescribed formula out of the aggregate cash flows from the underlying mortgages.

A wide variety of methods can be used to allocate the cash flows from the underlying mortgages, so there are numerous variations on the CMO theme. The typical structure involves each tranche receiving some share of the interest payments received but the principal payments are allocated sequentially, first to Tranche A, then to Tranche B, etc. In this case, Tranche A would receive all of the principal payments, including any prepayments, until the outstanding balance of the Tranche A was reduced to zero. Then, all future principal payments would go to the Tranche B until that tranche is fully repaid, and so on. Investors in each tranche receive their stated coupon rate applied to their outstanding balance on each coupon payment date.

Structural Diagram

It may be helpful to visualize the key steps in the creation of a CMO, which are shown in Figure 1 on the facing page and described as follows:

FIGURE 1. Mortgage Backed Securitization



Step 1 Bank issues mortgages to (10) homeowners, providing them with the cash they need to purchase the home. The bank will receive interest payments over 30 years and principal payments gradually such that the monthly payments are the same for the entire 360 month period. The bank is exposed to the risk that mortgage principal and/or interest are not repaid.

Step 2 Bank sells its mortgages to a third party who will later securitize the mortgages. In this case, we assume that the mortgages are all conforming mortgages and that Fannie Mae or Freddie Mac serve this role. At this point, the bank is no longer exposed to any risk. However, they may continue to service the mortgages — receiving the monthly payments and passing them along to Fannie Mae or Freddie Mac.

Step 3 Fannie Mae or Freddie Mac sells the mortgages to a Special Purpose Vehicle (SPV) but also provide a guarantee of the principal and interest payments. The SPV now receives the interest and principal payments either as scheduled or, if someone prepays the mortgage or defaults, they may receive the principal payments sooner than scheduled and then will no longer receive the interest payments. The SPV is not exposed to credit risk, but is exposed to the prepayment risk.

Step 4 SPV gets its funds to purchase the collateral by issuing bonds that pay interest and principal according to whatever schedule works given that their only source of funds are the cash flows from the mortgages (and the guarantee).

Step 5 Investors receive interest and principal payments only from the SPV's cash flows and therefore are exposed to the prepayment risk retained by the SPV. Priority of payments, and therefore the magnitude of prepayment risk exposure, can differ based on investor tranche.

Numerical Example — CMO Cash Flows with Three Tranches

Consider again the pool of mortgages described above containing \$1,000,000 worth of mortgages across 10 different borrowers. Now assume that we create a CMO with three Tranches A, B, C. Tranche A has a total value of \$100,000, Tranche B \$400,000 and Tranche C \$500,000. Further assume that each tranche earns interest monthly at an annual rate of 8% (the same rate as the underlying mortgages) and that scheduled and unscheduled principal payments are paid to Tranches A, B and C *sequentially*.

We will again determine the scheduled and actual cash flows in month 3 for the three different tranches assuming that one of the mortgages prepays at the end of month 3.

In this CMO, each class receives interest payments on a monthly basis equal to the 8% annual rate multiplied by their outstanding balance. They also receive their sequentially ordered principal payments. As a result, all principal payments whether scheduled or unscheduled go to Tranche A until that tranche is fully repaid, then all subsequent principal payments go to Tranche B until that tranche is fully repaid, and then finally Tranche C receives all remaining principal payments.

Consider the scenario where principal payments occur as planned. The pool will receive interest payments from the underlying mortgages which are used to make the interest payments to the various tranches based on the coupon rate for the tranche at the outstanding principal balance at the start of each period for the respective tranches. So to properly allocate the cash flows, we have to track the total interest payments and the total principal payments received on the underlying mortgages, as well as the outstanding balance of the CMO tranches.

Table 2 summarizes the key information we need from the underlying mortgages.

TABLE 2. Cash Flows from Underlying Mortgages (Payments as Planned)

	Each Mortgage		All Mortgages		
Month	Interest	Scheduled Principal	Interest	Scheduled Principal	Unscheduled Prepayment
1	667	67	6,667	671	0
2	666	68	6,662	675	0
3	666	68	6,658	680	0

From this, we can determine the principal payments made to each tranche as shown in Table 3.

TABLE 3. Principal Payments for Tranches A, B and C

	Mortgage	Principal Payments			Outstanding Balances (EOM)		
Month	Principal Paid	Tranche A	Tranche B	Tranche C	Tranche A	Tranche B	Tranche C
0					100,000	400,000	500,000
1	671	671	0	0	99,329	400,000	500,000
2	675	675	0	0	98,654	400,000	500,000
3	680	680	0	0	97,974	400,000	500,000

The principal payments in this case all go to Tranche A until A's balance is fully repaid. Therefore, the total cash flows in month 3 to Tranches A, B and C are scheduled to be \$1,338, \$2,667 and \$3,333.

And then once we know the beginning balances each month we can determine the interest payments to be made to the various tranches. In each period, the tranche interest payments equal the stated monthly rate times the outstanding balance at the beginning of the period. For example, in month 3 Tranche A's interest payments are $(8\%/12)(98{,}654) = \$658$. The calculations are similar for the other tranches, as shown in Table 4.

TABLE 4. Tranche Interest Payments

	Interest Payments					
Month	Tranche A	Tranche B	Tranche C	Total	Available	Shortfall
0						
1	667	2,667	3,333	6,667	6,667	0
2	662	2,667	3,333	6,662	6,662	0
3	658	2,667	3,333	6,658	6,658	0

Of course, we need to ensure that there is sufficient cash flow from the underlying mortgages to make the necessary interest payments to the three tranches. In practice this will likely lead to different coupon rates on the bond tranches that will reflect their relative risks, the impact of underlying defaults, etc.

Effect of Prepayment of Underlying Mortgage

Assume that at the end of month three one of the borrowers unexpectedly prepays the full outstanding balance of the mortgage. In this case, note that for any single mortgage, the initial balance was \$100,000 and in the first three months a total of \$203 in scheduled principal was paid for that particular mortgage. But then an additional $\$100{,}000 - \$203 = \$99{,}797$ is also paid (as shown in Table 5 on the following page). All of this payment goes sequentially to the tranches, as shown in Table 6 on the next page.

The Tranche A balance at the beginning of month 3 is only $98,654. This is less than the principal received that month, so Tranche A is fully repaid and the remaining principal goes to repaying Tranche B.

TABLE 5. Cash Flows from Underlying Mortgages (Prepayment in Month 3)

	Each Mortgage		All Mortgages		
Month	Interest	Scheduled Principal	Interest	Scheduled Principal	Unscheduled Prepayment
1	667	67	6,667	671	0
2	666	68	6,662	675	0
3	666	68	6,658	680	99,797

TABLE 6. Principal Payments for Tranches A, B and C

	Mortgage	Principal Payments			Outstanding Balances (EOM)		
Month	Principal Paid	Tranche A	Tranche B	Tranche C	Tranche A	Tranche B	Tranche C
0					100,000	400,000	500,000
1	671	671	0	0	99,329	400,000	500,000
2	675	675	0	0	98,654	400,000	500,000
3	100,477	98,654	1,824	0	0	398,176	500,000

The total cash flows to Tranches A, B and C are then $99,312 (their interest for the period, their scheduled principal from all 10 borrowers and the prepayment amount up to the total size of the tranche), $4,491and $3,333.

Because all of the principal payments go first to Tranche A, this effectively allocates significantly more of the prepayment risk to the Tranche A investors than to the Tranche B or Tranche C investors. As such, the A Tranche is likely to receive a larger share of the interest payments (a higher coupon rate) to compensate for the increased risk. All that matters is that there is sufficient interest overall from the mortgages to pay the interest to all three tranches.

Stripped MBS (IO & PO)

This is a more extreme case of a CMO in which certain investors receive their cash flows only from the interest payments that are made to the pool (the Interest Only or IO strip) and some receive their payments only from the principal payments that are made to the pool (the Principal Only or PO strip).

To understand the complex prepayment risk associated with an IO or PO, think about an investor who buys a PO for a pool with a single 30 year mortgage in it. To keep it simple, assume that this mortgage pays interest monthly but the principal is all paid at maturity in 30 years — or sooner if there is a prepayment.

Assume the investor expects the \$100,000 in principal, not scheduled until year 30, will actually be repaid in 10 years (that's about how long a typical homeowner stays in a house). If interest rates are 10%, the PO might be worth:

$$\frac{\$100{,}000}{1.1^{10}} = \$38{,}554$$

Remember that it doesn't receive any interest payments or interim cash flows, so this is just a zero coupon bond with an expected cash flow of \$100,000 in year 10.

Suppose the person actually stays in the house for the full 30 years, then our investor will have paid \$38,554 now and will receive the \$100,000 in 30 years. That gives him an actual annual return of:

$$\left(\frac{\$100{,}000}{\$38{,}554}\right)^{1/30} - 1 = 3.23\%$$

But if the person repays in the first year, the investor's actual return will be:

$$\frac{\$100{,}000}{\$38{,}554} - 1 = 159\%$$

You could do something similar for IOs. Notice that if you invest in an IO, you only receive the interest payments. If the person prepays the principal they stop making interest payments, so all of your cash flows just stop. This causes the value of the IO to drop substantially. On the other hand, if they stay longer than expected, more interest payments are made and the value of IO rises substantially. Unlike the PO investor who wants fast prepayments, the IO investor wants very slow prepayments.

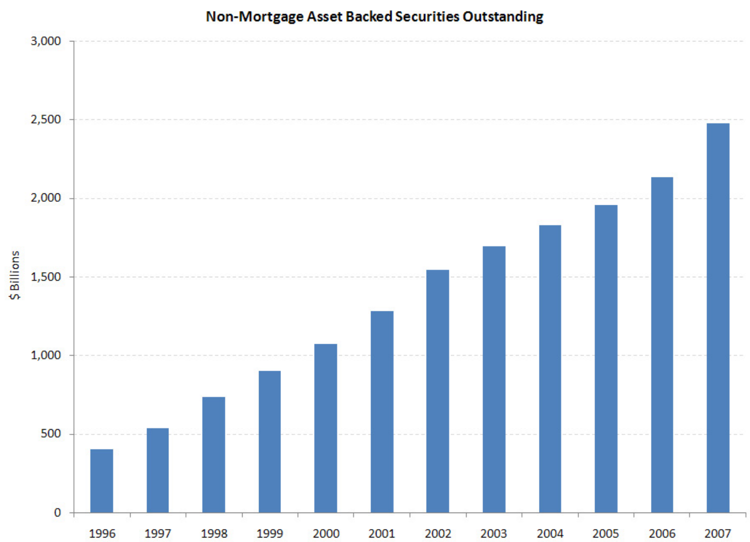
As we will see in Chapter 16 of BKM, duration measures the sensitivity of bond values to changes in interest rates. Usually, when interest rates rise, the value of a bond falls. It is interesting to note though that the value of an IO tends to move in the same direction as the interest rate — resulting in what is referred to as *negative duration*. When rates rise, prepayments slow (fewer people refinance) and the IO investor will receive more cash flows because more interest payments are made. This offsets the price decline from discounting at a higher rate. Similarly, when rates fall, the PV of the cash flows is higher, but the amount of the cash flows declines. The net effect is that the price rises when rates rise and falls when rates fall. Therefore, IOs are a useful tool for modifying the duration of a portfolio.

Asset-Backed Securitizations

The previous example focused on securitization of mortgages, where home mortgages served as *collateral* for bonds issued by the SPV. Mortgages work well because they have a steady stream of cash flows that can be sliced and diced to provide different cash flow and risk profiles for different investors. But certainly any other asset that produces cash flows can be used as collateral, including auto loans, credit card receivables, home equity loans, student loans, etc.

Figure 2 on the following page shows the growth of these markets from 1996-2007 (more detail by category is shown in BKM Figure 1.2).

FIGURE 2. ABS Outstanding 1996–2007 (Source: www.sifma.org)



Subprime Mortgage Securitization

The example shown earlier for a CMO made the simplifying assumption that a creditworthy counterparty (Fannie Mae or Freddie Mac) provided a guarantee of the mortgage payments. This guarantee could have come from other parties besides Fannie Mae or Freddie Mac, or it could simply be passed along to the investors.

In recent years, banks became very aggressive in providing mortgage loans to borrowers with very low credit quality (subprime borrowers) or did so with little or no effort to verify stated income, assets or ability to repay the mortgage. They tended to do so because they were able to still sell the mortgages into SPV's that would securitize the mortgage risk — both the prepayment risk as discussed above and the credit risk as well.

In theory, there should be nothing inherently wrong with this process, as investors are free to make their own risk assessment and simply demand that they be fairly compensated for assuming this risk. If the mortgage rates paid by borrowers with high probabilities of defaulting are high enough, and other forms of security such as the value of the home are sufficient to eliminate large losses in the event of default, then it should be possible for investors to be appropriately compensated for the risk they are assuming. Some in the lower tranches will be taking the *first loss* risk and others in the higher tranches will be taking more remote risk that will require a large number of defaults in an otherwise diversified pool.

On the surface, this should be no different than other asset-backed securitizations discussed above. But subprime mortgage securitizations were indeed different — in a very important way.

Why Were Subprime Mortgage Securitizations Different?

As subprime mortgage origination grew ($1.2 trillion was originated in 2005 and 2006 alone), the borrowers were *not* paying higher mortgage rates than more credit-worthy borrowers. In fact, they often paid lower rates initially either because of teaser rates that would apply for the first few years before being reset to much higher risk-adjusted rates or because of features that allowed some or all of the monthly payments to be deferred and rolled into the principal balance.

What this meant in practice was that *subprime mortgage securitizations depended heavily on housing prices increasing*. If the house price increased before the end of the teaser rate period, subprime borrowers would normally be able to refinance their mortgages and start over rather than have to actually pay the higher mortgage rates which most of them clearly could not afford. This would be treated in the securitization as a prepayment and the various investors would be repaid before the borrowers' inability to actually make their payments resulted in a default. Similarly, so long as house prices increased, an actual default would potentially lead to minimal net losses for investors so long as they were able to foreclose on the house and sell it for close to the outstanding loan balance.

When the housing bubble finally burst, prices began to fall and subprime borrowers began to default in huge proportions beginning in 2007, it became clear that the investors were not carefully assessing the risk and likely were not being fairly compensated for the risk they assumed. In addition to the fact that these investors were inherently making a huge bet that house prices would rise, many relied solely on rating agencies to assess the risk. The rating agencies proved to be less than diligent in their own risk assessment.

One problem with the rating agencies' analyses was that it relied too much on recent history, which reflected a relatively robust economy and declining interest rates. Another is that it relied on mortgage default experience from a period of rapidly rising home prices, which we saw before essentially precluded the need for an actual default due to the ability to refinance or sell the home. And third, there is some evidence that in certain cases they may have underestimated the risks intentionally due to significant conflicts of interests arising from the fact that they were paid by the issuers rather than the investors.

The sharp and simultaneous decline in the U.S. economy and home prices across the country caused a large spike in both defaults and losses on subprime mortgage securitizations. The hardest to be hit were the senior-most tranches whose risk was originally thought to be extremely remote. Although by the end of 2008 most of those senior tranches had not actually experienced default losses, their mark-to-market valuations were substantially below the par value given the greatly increased expectation that losses would be incurred and the enormous risk margins investor suddenly demanded for those large, illiquid and now obviously risky tranches.

This inherent, and substantial, bet on house price appreciation made subprime securitizations unique. As will be discussed below, the fact that various tranches of subprime securitizations made their way into collateralized debt obligations (CDOs) made the effect of this substantially worse for the entire financial system than it might otherwise have been.

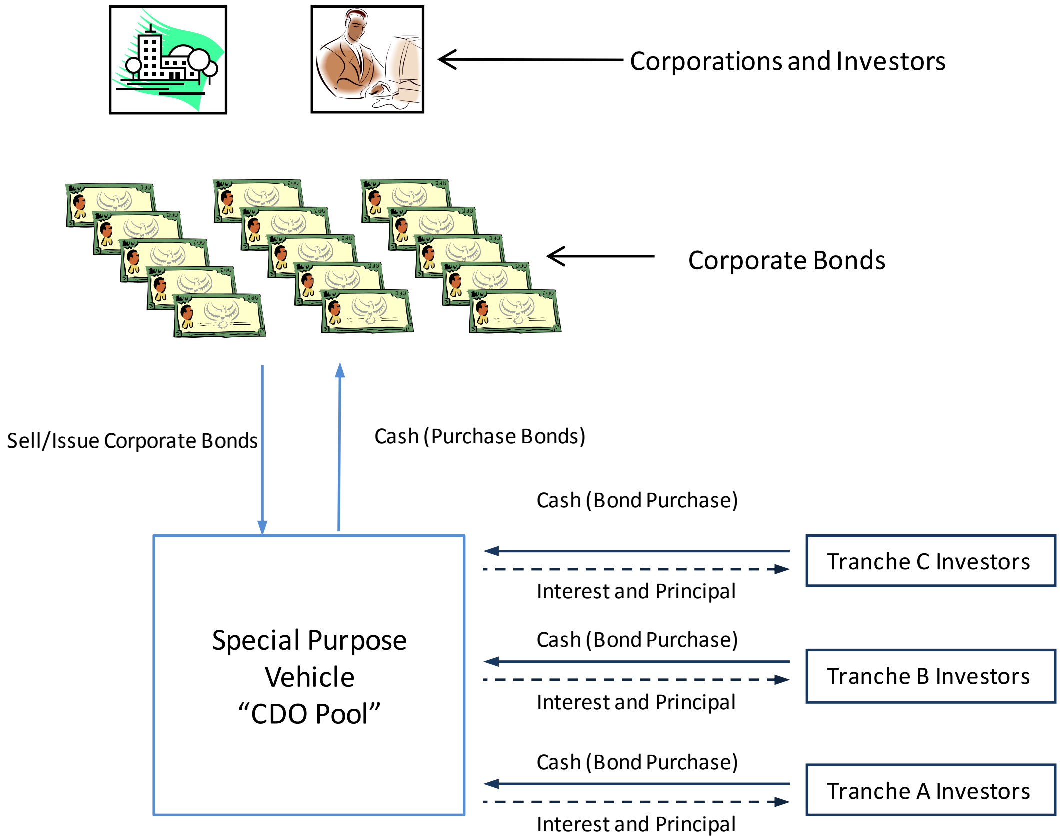
Collateralized Debt Obligations (CDOs)

In simple terms, CDOs are identical to the CMOs we've been discussing except that the collateral in the SPV is a generic bond (a "debt obligation" rather than a mortgage *per se*).

When first developed, the CDOs primarily contained corporate bonds of different credit qualities. Pooling, for instance, a large number of non-investment grade bonds from diverse counterparties could in theory make it possible to create high-risk "first loss" tranches and more remote "senior" or "super senior" tranches for which the respective investors would demand different levels of compensation. So even a pool of low-rated bonds could be used to manufacture a tranche of relatively high rated bonds for which losses could only be realized if substantially all of the bonds defaulted (whereas the first loss tranches would suffer losses if even a single bond defaulted). An appropriately diversified pool of corporate bonds might make the probability that substantially all of the bonds defaulted very low, allowing the senior-most tranches to be viewed as relatively risk remote.

A diagram of the key elements of the structure looks a lot like that of a CMO, but for the fact that the collateral in the pool consists of corporate bonds, either bonds newly issued by corporations or existing bonds purchased from investors. This is shown in Figure 3 on the next page.

FIGURE 3. Collateralized Debt Obligations

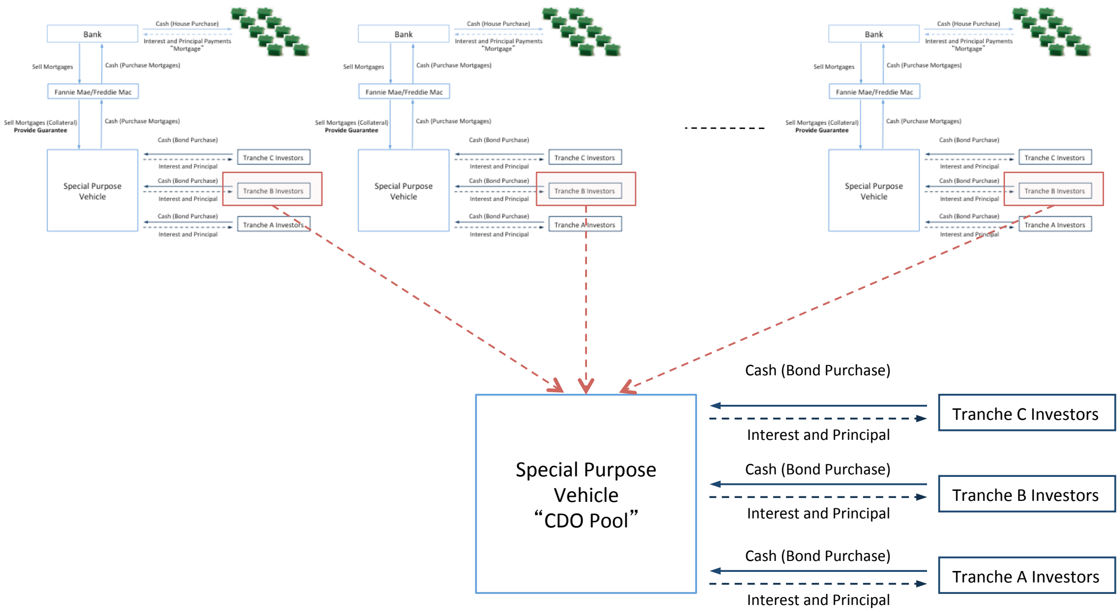


This innovation led to enormous growth in the CDO market that far outstripped the size of the actual bond market in part because CDOs could be created with *synthetic* bonds that were really just derivatives tied to the performance of actual corporate bonds (such as credit default swaps).

In addition, risky tranches of the kinds of mortgage-backed, asset-backed and subprime securitizations described earlier could be used as the "bonds" in the CDO pools. These various securitization tranches would then be pooled and tranched again, with investors in the resulting CDO tranches taking on increasingly more complex risk to analyze and monitor.

Figure 4 on the following page shows the mezzanine tranches of asset-backed securities being pooled into a CDO.

FIGURE 4. CDO of ABS



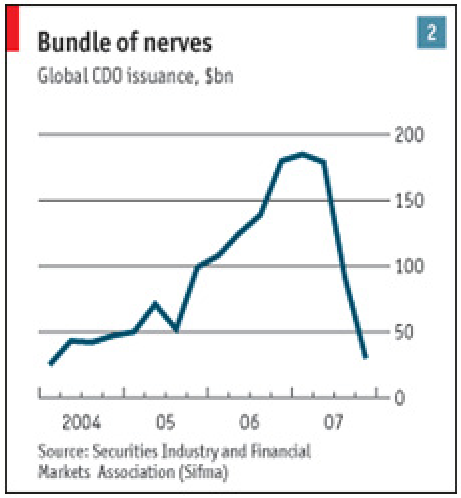
This piling on of more and more complex structures eventually reached a critical tipping point in 2007 and played a critical role in the collapse of the balance sheets of several global financial institutions once defaults on the mortgage assets began and risk across a wide range of asset classes was repriced.

First, CDO's that contained subprime mortgage securitizations began to experience substantial defaults and caused huge losses for those holding the most junior tranches. Second, the losses on the junior tranches made the senior tranches more exposed to potential losses and caused their mark-to-market valuation to plummet substantially. It is these losses on the senior-most tranches that resulted in hundreds of billions of dollars of write-downs for global financial institutions who mistakenly thought that "remote" risk meant "no risk". And third, the sudden need for investors everywhere to reduce leverage and reduce risk meant that certain CDO tranches became impossible to sell to investors.

As a result, the volume of CDO issuance dropped dramatically, as shown in Figure 5 on the next page.

The text also notes that many banks used Structured Investment Vehicles (SIVs) to raise short-term funds in order to purchase other forms of longer-term, riskier debt — in many instances tranches of very risky CDOs. These borrowings were secured by the longer-term and riskier assets. During the financial crisis that began in 2007, short-term funding became scarce (especially if it was backed by risky assets) and many banks were unable to roll over their short-term debt to fund their longer-term and increasingly illiquid assets — at least not at the rates that

FIGURE 5. CDO Issuance 2004–2007



they had previously counted on paying for that debt. These SIV's had to be unwound and the banks had to retake possession of the assets.

Finally, the rating agencies came under intense pressure during all of this, since they somewhat recklessly, and with clear conflicts of interest, gave much higher credit ratings to CDO tranches than some believe were ever justified.

In the Coval, Jurek and Stafford reading we will examine this issue in more detail. There, we will show that errors in estimates of the default risks for the underlying mortgages can lead to erroneously high ratings for senior tranches with high probabilities of default. We will also examine how the process of securitization can create senior tranches that contain (catastophically) high levels of systematic risk. As discussed elsewhere, investors should demand high risk premiums for assuming systematic risk. Failure to recognize the level of systematic risk in senior securitization tranches will therefore lead to substantial mispricing of those tranches regardless of the accuracy of the default probability estimates.

In the next reading and in Hull Chapter 24 we will discuss the role that incorrect correlation assumptions by the rating agencies contributed to their willingness to assign AAA ratings to senior and super-senior CDO tranches that in fact had substantial probabilities of default at issuance.

Coval, Jurek & Stafford: The Economics of Structured Finance

Introduction

In this reading the authors examine two aspects of securitizations that contributed significantly to the financial crisis of 2007-2009:

- Ratings for senior tranches of securitizations are extremely sensitive to modest imprecision in evaluating the default probabilities and default correlations among the underlying risks.
- Securitized tranches contain significant exposure to systematic risks.

Manufacturing AAA-rated Securities

Of all of the corporate debt issued globally, less than one percent of it was rated AAA prior to 2007, which was simply not sufficient to meet all of the demand for such bonds coming from institutional investors looking for investment opportunities or institutions looking to borrow cheaply by being able to pledge highly rated collateral. This created enormous incentives to *manufacture* highly rated debt securities out of existing bonds.

One way this was done was via a cash CDO, which was described in the previous reading. To emphasize here the *rating* implications of structuring a CDO, consider a simple two bond example where each bond has a \$1 face value and 10% probability of default. Further assume the default risk for each is independent (zero default correlation).

Suppose we created a \$2 notional value CDO by issuing a junior tranche of \$1 that would absorb the first loss and a \$1 senior tranche to absorb the second loss. Even though the probability of default for each bond is 10%, and hence each of their credit ratings would be below investment grade, the probability of *both* bonds defaulting is only $10\% * 10\% = 1\%$ and so the rating for the senior tranche would be quite high. In essence, we were able to take a market that consisted solely of low-rated bonds and manufacture a single bond tranche with a very high rating. Of course, now that the junior tranche experiences losses if either of the bonds defaults, its default probability is higher than either of the underlying bonds (and is equal to $1 - .9^2 = 19\%$.).

Note that had the default risk been perfectly correlated, then either none of the bonds defaults or both bonds default. In that case, the probability of both defaulting is the same as the probability of either default, or 10%. This means that the senior tranche would have the same rating as the underlying bonds and there would be no way to manufacture a highly rated tranche.

This problem can be resolved therefore by using a large number of imperfectly correlated bonds. As the number of bonds increases and the average correlation among the bonds decreases, the portion of the combined structure that could achieve a high rating increases.

And as discussed in the previous reading, another way to increase the notional value of highly-rated tranches is to combine a large number of junior tranches created through the CDO process used above into a new pool, out of which highly rated tranches can be created. In all cases, risk is not disappearing, it is just being pushed down to the junior tranches, leaving less risk in the senior tranches.

The Challenge of Rating Structured Finance Assets

It is important to appreciate that, given the complexity of modeling the default risk for large pools of corporate bonds — and even more so for pools of tranches of MBS, ABS or other CDOs — and the dependence on the correlation assumptions across the underlying bonds, trillions of dollars of CDO tranches were purchased by investors based, in many cases, entirely on the rating assigned to it by the rating agencies (S&P, Moody's and Fitch). Given this, it's somewhat stunning how imprecise this process was.

Recall that the rating agencies, prior to 1997 or so, focused almost entirely on rating single-issuer corporate bonds where the risks of default were driven by business conditions and which were assessed through analysis of financial statements. In addition, this process never had to consider the correlation across issuers. Once that rating agencies began to rate CDOs, however, their historical skill set was somewhat irrelevant and entirely new methods and models were needed.

In addition, small errors in the risk assessment of the underlying assets are magnified when these are tranched in CDOs, and further tranched in CDO-squareds.

Sensitivity to Parameter Estimates

To demonstrate the sensitivity, the authors create a simulation model of CDO pools consisting of 100 bonds, each of which has a 5-year default probability of 5% and a recovery rate of 50% of the notional amount. They use a one-factor Gaussian copula model (as discussed in Hull Chapter 23) for the default on each bond in the pool and use a factor sensitivity of 20% for each bond[1], which results in the correlation for any pair of bonds is the same within the same pool (bonds in different pools are independent).

They then create CDO tranches with cutoffs of 6% for the junior tranche (that is, the junior tranche absorbs the first 6% of losses in the portfolio) and 12% for the mezzanine tranche (the mezzanine tranche absorbs losses from 6% - 12%). The senior tranche absorbs losses greater than 12%.

Finally, they use the mezzanine tranches from 40 independent CDOs (each one created using the same assumptions as outlined above) to create a CDO-squared and apply the same cutoffs for the three tranches of this CDO-squared.

[1]In the paper, the authors refer to the *correlation* as being 20%. However, their model actually used a factor sensitivity of a_i = 20%, so the correlation in their baseline model was really 4%. Similarly, the sensitivity tests they performed varied this a_i parameter rather than the correlation.

After simulating the loss experience for each of the tranches they summarize the probability of default and the expected payoff (as a percent of the notional value) for each tranche using their baseline assumptions. By comparing the default probability of each tranche to historical default rates for bonds, they approximated the credit rating each tranche would receive. Table 1 shows the results of the authors' simulations under the baseline assumptions.

TABLE 1. Baseline Simulation Results

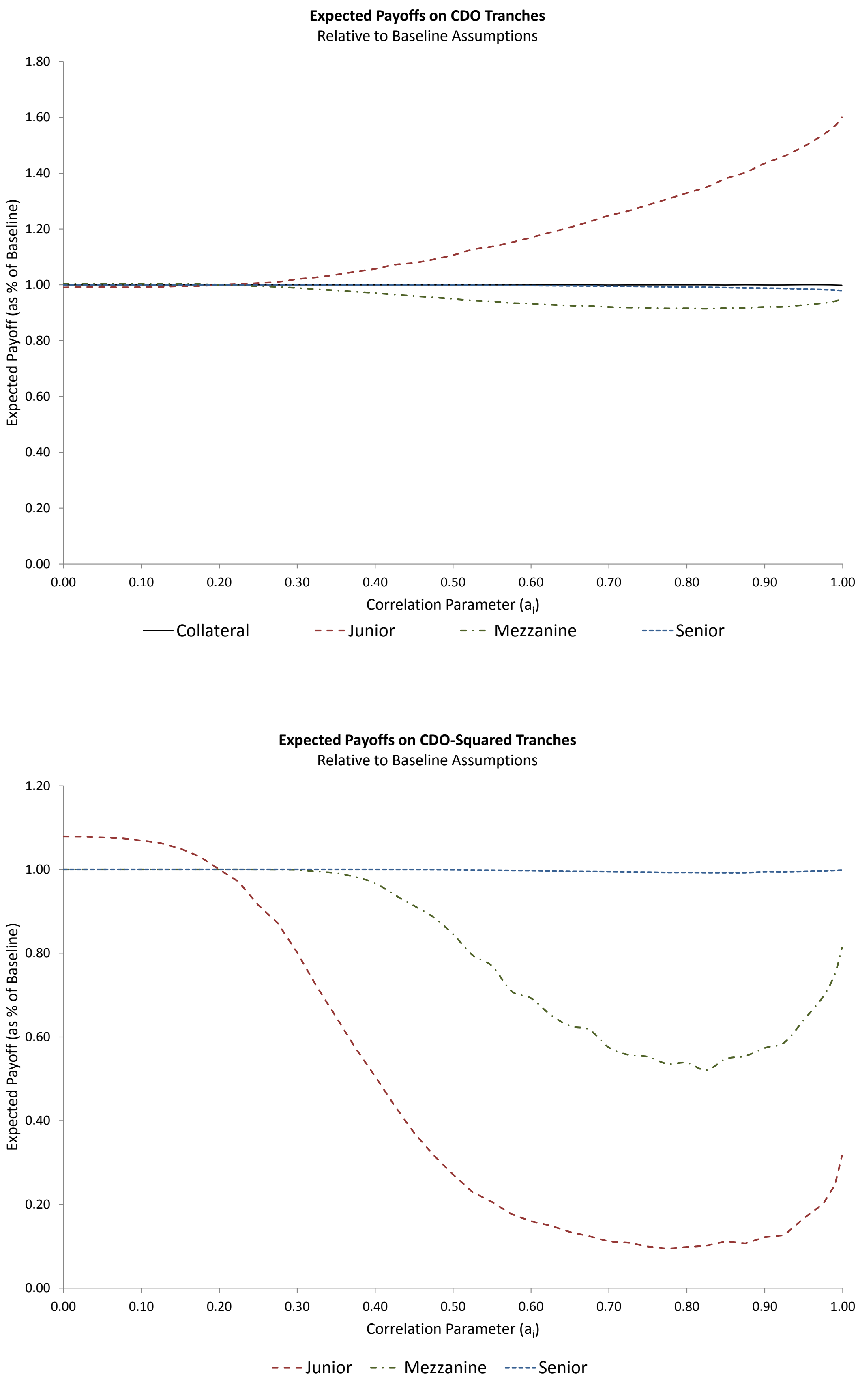
	Attachment Points	Default Probability	Expected Payoff	Rating
CDO				
Junior	0% - 6%	0.9752	0.59	NR
Mezzanine	6% - 12%	0.0207	>0.99	BBB
Senior	12% - 100%	< 0.00%	>0.99	AAA
CDO-Squared				
Junior	0% - 6%	0.5694	0.93	C
Mezzanine	6% - 12%	< 0.00%	>0.99	AAA
Senior	12% - 100%	< 0.00%	>0.99	AAA

Table 2 and Figure 1 on the following page show how sensitive the expected payoff is to the correlation parameter assumption.

TABLE 2. Sensitivity of Results to Correlation Assumption

	CDO					
	Junior		Mezzanine		Senior	
Correlation	Payoff	% of Baseline	Payoff	% of Baseline	Payoff	% of Baseline
0.00	0.58	0.99	1.00	1.00	1.00	1.00
0.10	0.58	0.99	1.00	1.00	1.00	1.00
0.20	0.59	1.00	1.00	1.00	1.00	1.00
0.50	0.65	1.11	0.95	0.95	1.00	1.00
0.90	0.84	1.44	0.92	0.92	0.99	0.99
1.00	0.94	1.60	0.95	0.95	0.98	0.98
	CDO Squared					
	Junior		Mezzanine		Senior	
Correlation	Payoff	% of Baseline	Payoff	% of Baseline	Payoff	% of Baseline
0.00	1.00	1.08	1.00	1.00	1.00	1.00
0.10	0.99	1.07	1.00	1.00	1.00	1.00
0.20	0.92	1.00	1.00	1.00	1.00	1.00
0.50	0.25	0.27	0.85	0.85	1.00	1.00
0.90	0.11	0.12	0.57	0.57	0.99	0.99
1.00	0.29	0.32	0.81	0.81	1.00	1.00

FIGURE 1. Sensitivity to Correlation Parameter



Notice that for the basic CDO, the junior tranche expected payoffs increase as the correlation increases. When the correlation approaches 1.0, the expected payoffs on each tranche have to approach the expected recovery of the underlying bonds since at that point each of the 40 bonds either defaults or doesn't at the same time. While the expected payoffs for the mezzanine and senior tranches decline as the correlation increases, the effect is small for the mezzanine tranche and insignificant for the senior tranche. That is, except under extreme assumptions, the senior tranche expected cash flows are relatively stable.

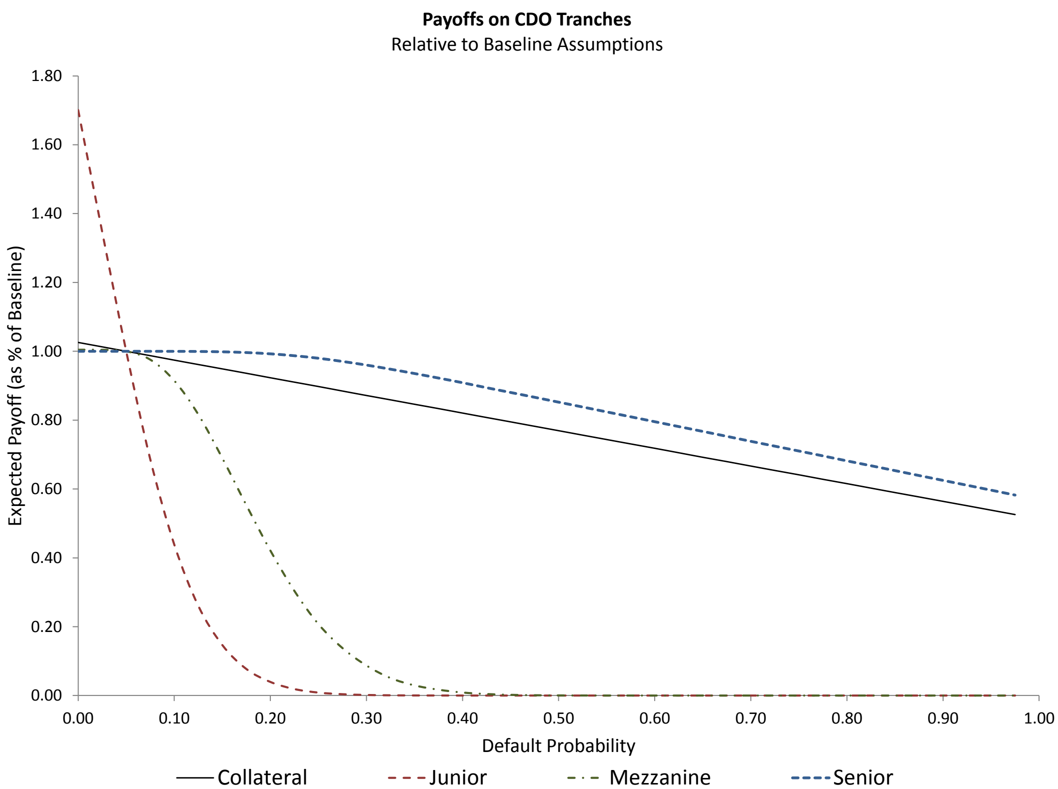
However, for the CDO-squared, the junior and mezzanine tranches show extreme sensitivity to the correlation parameter assumption, with expected payoffs declining sharply until the correlation parameter is extremely large (above 80%). At that point, the payoffs begin to increase as more of the risk is shifted to the senior tranche.

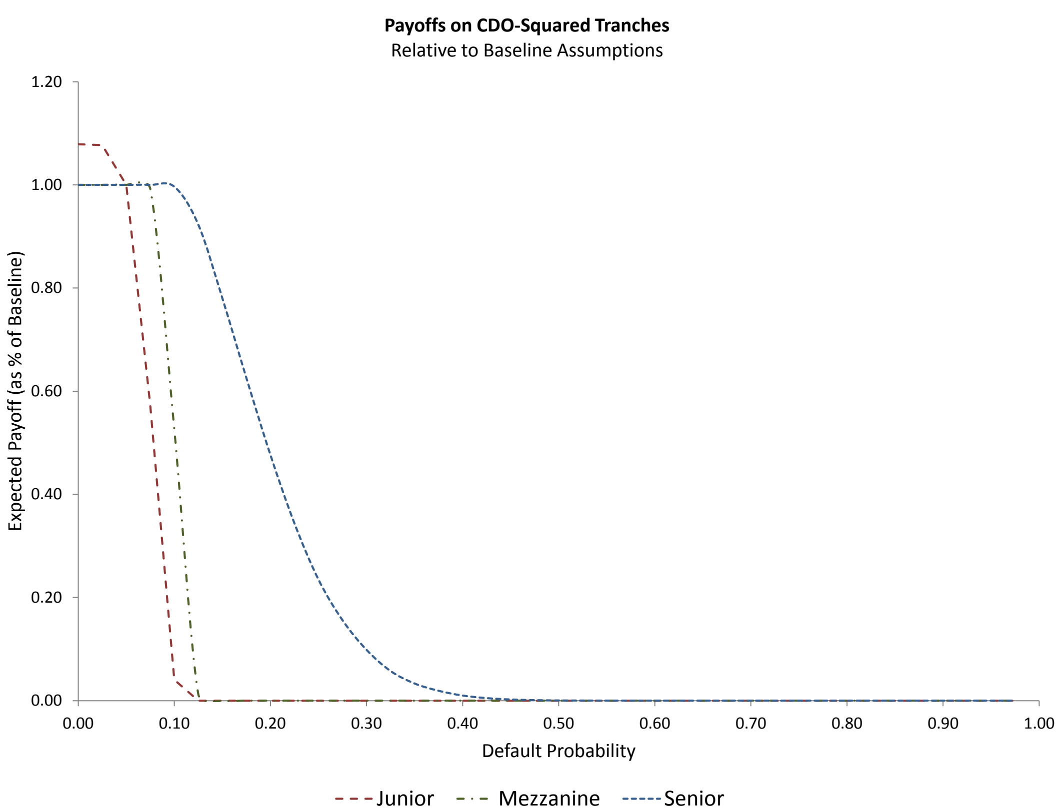
More striking is how sensitive the tranches are to the assumed default probabilities for the underlying bonds. Even modest errors in estimating the default probability can result in dramatic decreases in expected payoffs for the CDOs and even more dramatic impacts on the expected payoffs for the CDO-squared. Table 3 and Figure 2 on the following page show the expected payoff for each tranche under different assumptions for the default probability of each bond.

TABLE 3. Sensitivity of Results to Default Assumption

CDO						
	Junior		Mezzanine		Senior	
Default Prob	Payoff	% of Baseline	Payoff	% of Baseline	Payoff	% of Baseline
0.00	1.00	1.70	1.00	1.00	1.00	1.00
0.05	0.59	1.00	1.00	1.00	1.00	1.00
0.10	0.26	0.44	0.91	0.91	1.00	1.00
0.15	0.09	0.15	0.69	0.69	1.00	1.00
0.25	0.01	0.01	0.21	0.21	0.98	0.98
0.50	0.00	0.00	0.00	0.00	0.85	0.85
CDO Squared						
	Junior		Mezzanine		Senior	
Default Prob	Payoff	% of Baseline	Payoff	% of Baseline	Payoff	% of Baseline
0.00	1.00	1.08	1.00	1.00	1.00	1.00
0.05	0.93	1.00	1.00	1.00	1.00	1.00
0.10	0.04	0.04	0.53	0.53	1.00	1.00
0.15	0.00	0.00	0.00	0.00	0.78	0.78
0.25	0.00	0.00	0.00	0.00	0.23	0.23
0.50	0.00	0.00	0.00	0.00	0.00	0.00

FIGURE 2. Sensitivity to Default Probability Assumption





Are the AAA-rated Tranches Safe?

Be careful when analyzing the apparent stability of the senior tranches, especially for the CDO-squared.

Recall that the authors still assumed that the bond defaults across the different pools were independent. To the extent that defaults across the universe were driven by a common factor, even the senior tranches would show dramatic sensitivity. Further, those tranches tended to have very small risk margins, so even modest changes in the expected payoffs could result in massive losses. And finally, notice that while their expected payoffs don't appear to be very sensitive, the AAA rating is very sensitive and the resulting repricing when the rating falls to say BBB- would be extreme.

But there is another, much more important caveat to keep in mind. The AAA-rated tranches may have relatively remote risk, but they are extremely sensitive to the impact of systematic risks. As discussed in the BKM readings, the pricing of these tranches should be a function of not only the probability of default and the expected payoffs, but also of the degree to which the losses occur at a particularly good or bad time for the investor. It turns out that AAA-rated tranches of securitizations, and particularly re-securitizations as in CDO-squareds, indeed carry substantial systematic risk. They are akin to what the authors, in a companion paper, referred to as *economic catastrophe bonds*. As such, while their default risk was low, the prices should have reflected the nature of the risk they contained.

CDO and CDO-squared Sensitivity and Subprime Mortgages

Consider the events of 2007-2009 in the context of the sensitivity analysis above.

Notice *all* subprime mortgages had default probabilities that were very sensitive to home price appreciation since without rising home prices these borrowers who never really could have afforded their mortgages to begin with, would not be able to refinance. As a result, CDOs containing subprime mortgages and, to an even greater extent CDO-squareds that contained mezzanine tranches of subprime CDOs, were dependent upon default probabilities that may not have accurately reflected the risks of a nationwide recession and/or the effects of a nationwide decline in home prices. Further, since these two effects would have impacted all subprime mortgages, it is very likely that the default correlations were wildly underestimated as well.

Pricing of Systematic Risk

As discussed above, the highly-rated tranches of CDOs and CDO-squareds contained highly concentrated systematic risk. Unfortunately, the rating agencies' analysis did not capture this and implicitly assumed that holding a AAA-rated securitization tranche was akin to holding a similarly rated corporate bond. Even if one thought that the rating agencies' analyses were accurate (see below), the pricing of these tranches should have reflected substantially higher risk margins.

Why Did the Structured Finance Market Grow So Spectacularly?

A number of important factors contributed to the spectacular ten-fold growth in the structured finance market in under a decade:

1. Seemingly Attractive Yields — Because the rating agencies assigned AAA-ratings to substantial portions of the notional value of the structured finance tranches and the tranches paid marginally higher yields than comparably rated corporate bonds, investors thought they were getting attractive risk-adjusted returns. In fact, the ratings reflected biased historical default experience and the yields failed to reflect the degree of systematic risk assumed.

2. Extreme Market Optimism — It is hard to overemphasize how many market participants shared the overly optimistic assumptions used by the rating agencies with regard to subprime mortgage defaults. Many knowledgeable market participants simply didn't believe that home prices could decline and some analysts didn't even include the possibility of this in their models.

3. Too Little Appreciation for Fragility of Ratings — Consider the fact that bank regulatory capital requirements were tied to ratings and investors, in large part, outsourced their risk analysis to the rating agencies, without regard to the fact that the ratings themselves, as the authors showed, were very sensitive to the assumptions.

4. Perverse Incentives for Rating Agencies — Didn't anyone know that this was a house of cards? Well, perhaps some did but did not have the incentive to "stop dancing". For instance, rating agencies are paid by the issuers and not the investors, so they had strong incentives to continue to provide high credit ratings to new issuances and to effectively become part of the structuring team.

5. Perverse Incentives for Banks — Banks of course collected huge fees for creating the structured finance products and were happy to continue providing the product to investors. Interestingly though, the banks also faced very low capital requirements for holding onto the senior-most tranches. This motivated them to hold onto large portions of these tranches, effectively falling for their own sales pitch. When the market collapsed, it was the banks' holdings of senior structured finance claims that resulted in the massive losses reported as those holdings had to be marked to reflect their revised expected default rates, their revised ratings and the revised market risk premiums.

For additional insights, you can refer to Chapter 8 in the Eighth Edition of Hull for further discussion of the points made in this reading.

Practice Questions

Question 1. Suppose you had three bonds, each with \$1 in face value and a 10% probability of default. These three bonds will be combined into a CDO with \$3 in notional value and three separate tranches, each with \$1 in notional value. If the default probabilities are independent, what is the probability of default for the junior, mezzanine and the senior tranches?

Solution. Start with the senior tranche, which will suffer a default only if all three of the bonds defaults. Since their defaults are independent, the probability of this occurring is $p_s = .1^3 = .001$.

The mezzanine tranche default probability is harder but can be found from the following. First, the probability that any two tranches default is

$$p = \binom{3}{2}.1^2.9 = 3(.01)(.9) = 2.7\%$$

Since the mezzanine tranche will default if any two default or if all three default, the mezzanine default probability is $p_m = 2.8\%$

Finally, the junior tranche defaults if any of the bonds defaults and so it's default probability will be one minus the probability that none of them default, or $p_j = 1 - .9^3 = 27.1\%$.

Here, we were able to create a CDO in which two-thirds of the notional value has lower default risk than the underlying bonds, and the other third is a bond with a very high probability of default.

Question 2. When rating structured finance transactions, the rating agencies typically provided the same expected default rates and expected recovery statistics to investors and used those as the basis for their ratings. Why do Coval, Jurak and Stafford argue that this could have caused investors to significantly underprice the risks in structured finance tranches associated with subprime mortgages?

Solution. One point to note is that CDOs, and in particular CDOs of tranches of asset-backed securitizations or CDO-squareds, had default probabilities and expected cash flows that could be substantially more sensitive to assumptions than was often recognized. Any intentional or unintentional bias in the underlying asset default and recovery assumptions would have resulted in materially biased default and recovery rates for the securitized tranches.

Second, unlike other debt obligations, subprime mortgages had a unique feature in that they were closely tied to, and highly sensitive to, home price appreciation. Particularly for the so-called liar loans or the NINJA (no income, no job or assets) loans, in the absence of home price appreciation these mortgages were almost certain to default. This significantly impacted not only the reliability of default rate assumptions tied to recent historical default rates, but it also caused the underlying asset default rates to be much more highly correlated.

Finally, the effect of securitization and particularly re-securitization is to load the senior-most tranches with systematic risk. In this case, bonds that effectively serve as economic catastrophe bonds ought to be priced with substantial risk margins by most investors. Investors relying solely on ratings, which don't provide any indication of the degree of systematic risk, and yield spreads for more plain-vanilla bonds, would almost certainly underprice the risk associated with these tranches.

Question 3. List five reasons that the structured finance market grew so substantially in the years leading up to the financial crisis.

Solution. Five reasons given are:

1. Seemingly Attractive Yields
2. Extreme Market Optimism
3. Too Little Appreciation for Fragility of Ratings
4. Perverse Incentives for Rating Agencies
5. Perverse Incentives for Banks

Question 4. Coval, Jurak and Stafford argued that because the senior tranches of re-securitizations were more susceptible to parameter and model risk and contained primarily systematic risk as opposed to idiosyncratic risk, those senior tranches were substantially mispriced. What impact did this have on the ability of the structured finance market to grow so substantially?

Solution. Recall that the highly rated tranches of securitizations could account for the bulk of the notional value of a CDO. Nonetheless, the remaining junior tranches, while small, are somewhat "toxic" in that they have substantial default probabilities. Since someone has to be willing to buy these toxic junior tranches, underpricing the senior tranches allowed the issuers to overcompensate the junior tranches. Because of this there was robust demand for all tranches of CDOs, which allowed the machine to keep running. This fueled not only the incentive to keep packaging bonds, mortgages and other debt instruments into CDOs but also the motivation to continue to loan money to ever-riskier homeowners in order to create the underlying product that could be placed into the CDOs. This in turn fueled the housing bubble.

Question 5. What is it about structured finance bonds that make them so much more difficult to rate than corporate bonds?

Solution. Ratings on corporate bonds do not have to consider the correlation of defaults across different companies or different assets within a given pool. However, this is a critical aspect of the rating of various tranches of CDOs and other structured finance instruments.

In addition, structured finance tranches are rated in a way that makes them far more sensitive to the default probabilities assigned to the underlying bonds.

Question 6. You are considering purchasing a senior tranche of a CDO that has been assigned a AAA-rating from S&P to finance catastrophe risk exposure in an insurance policy your company has written. You notice that the yield on the bond is comparable to, but slightly higher than, the yield on other AAA-rated financial instruments such as commercial paper and bonds issued by AAA-rated companies. You are concerned about the value of the CDO at the time that you may have to sell it to pay a claim but have seen studies that show that the cash flows on senior tranches of CDOs are not very sensitive to changes in default and correlation assumptions. What are some reasons why you should be concerned about the appropriateness of the senior tranche for your portfolio?

Solution. There are several issues to be concerned with:

- Senior tranches of CDOs are difficult to rate and so it may be inappropriate to rely solely on S&P's analysis, especially because they are paid by the issuer for their ratings and this may give them some incentive to bias their ratings upwards.

- Ratings are highly dependent on correlation assumptions for the bonds in the pool. This is something that S&P may not have the ability to measure accurately, may change over time, etc.

- Although the cash flows of senior tranches of CDOs are not very sensitive to assumptions, except in extreme scenarios, their *ratings* are very sensitive to assumptions. Purchasing this tranche exposes you to potentially large swings in the price of the bond in the event the rating agency's assumption change and the bond is downgraded.

- The AAA-rating and the yields on other AAA-rated bonds doesn't tell you anything about the appropriate yield on the CDO tranche. The latter is far more sensitive to systematic risk and so it's yield spread relative to other AAA-rated debt should be much higher.

Question 7. The following table depicts the results of a simulation analysis of a CDO formed with 100 bonds:

TABLE 4. Baseline Simulation Results

	Attachment Points	Default Probability	Expected Payoff	Rating
CDO				
Junior	0% - 6%	0.9752	0.59	NR
Mezzanine	6% - 12%	0.0207	>0.99	BBB
Senior	12% - 100%	< 0.00%	>0.99	AAA

The model assumed the bonds in the pool had essentially independent default probabilities (the default correlation across the bonds was only 4%). If instead we assumed that there was a large degree of default correlation across the bonds, what would you expect to happen to the probability of default and the expected payoff (as a percent of the notional amount) for the junior tranche of this CDO?

Solution. The results show that the junior tranche contains most of the default risk, as evidenced for instance from the low expected payoff of only 59% of the notional value for the tranche. As the correlation increases towards 1.0, less of the default risk can be pushed down from the mezzanine and senior tranches to the junior tranche — all bonds will either default or not default at the same time and the default probabilities will be the same for all tranches. This means that the junior tranche default probabilities will fall and their expected payoffs will rise.

Hull Chapter 24: Credit Derivatives

Introduction

There are two elements of this reading that are relevant to the CAS exam. The reading begins with a discussion of various *credit derivative* products, the most important of which is the (single-name) credit default swap (CDS). It then provides an overview of (multi-name) CDOs. The latter discussion builds on the securitization background reading and the Hull Chapter 8 summary earlier in this section of the notes.

The rest of the chapter covers models used to estimate the cash flows (and values) for CDOs and other derivative securities. This material is not on the CAS syllabus, so I will not discuss it in these notes.

Credit Default Swaps

A credit default swap represents insurance against a default by a particular company on a particular issuance of debt. It involves periodic payments by the party purchasing the insurance and, in the event of default, a payment by the seller.

The payment made is usually the difference between the market price of the bond and par, but could also be handled by the payment of the par value from the seller in exchange for physical delivery of the bond.

Since it is possible (perhaps likely) that neither the buyer nor sell of the insurance has any real connection to the company or its bonds, some terminology used in these deals is important to keep track of:

- Reference Entity — This is the entity whose credit rating is being protected.
- Reference Obligation — The bond that is used to monitor whether a credit event occurs and to determine the swap payoff.
- Notional Principal — The par value of the reference obligation.
- Credit Event — This is the technical definition of what constitutes a "default", such as the failure to make a payment when it is due, a restructuring of debt or bankruptcy. It could also include the default of other bonds issued by the company besides the reference obligation (known as a cross-default).

Credit Default Swap Spread

The *CDS spread* is defined as the payment rate such that the present values of the buyer and seller payments are equal. In other words, it is the payment rate such that the value of the swap is zero at inception.

A CDS can be used to hedge the credit risk in a corporate bond, though not perfectly. If you buy the bond and at the same time buy protection via a CDS, you have a nearly risk free bond. As a result, the CDS spread should be approximately equal to the differences in the par yields for the corporate bond and the risk free bond. If the risk free rate is the LIBOR/swap rate, then this difference will equal the asset swap rate (see Hull Chapter 23).

An arbitrage argument can be used to show that the difference between the CDS spread and the asset swap spread, referred to as the CDS-bond basis, should be approximately zero.

Suppose the CDS-bond basis were negative. That means that you could buy the corporate bond, buy protection with the CDS for less than the spread on the bond and earn more than the risk free rate on a bond that is nearly risk free.

Valuation of Credit Default Swaps — Numerical Example

To value a CDS, we merely want to be able to determine the risk neutral present values of both sides of the swap and solve for the annual CDS spread that would set the value of each sides' payments equal to each other.

To see this clearly, consider the example used by Hull in Tables 24.1 through 24.4 of the text for a 5-year CDS on a bond with a 40% recovery rate in the event of default and conditional default probabilities of 2% per year.

The CDS spread, s, is paid annually, at the end of the year, and in the event of default the accrued portion of the spread is also paid (i.e. default halfway through the year requires a payment of one-half the spread).

In the event of default, the CDS seller pays the face value of the bond less the recovery amount. We will assume that default always occurs halfway through the year and that the risk free rate is 5%.

- **Step 1: Determine Unconditional Default Probabilities**

  Table 1 on the next page contains the unconditional default probabilities (as seen at time zero) and the corresponding survival probabilities for each of the next 5 years, following the calculations from the reading to convert the 2% annual conditional default probability to unconditional default probabilities.

- **Step 2: Determine Protection Buyer's Payments**

  Next, we need to calculate the expected CDS payments made by the protection buyer. Here, we actually need to do two calculations, one to reflect the annual spread that is paid assuming that default has not already occurred on each due date (this will use the survival probabilities at each possible payment date) and one to reflect the accrual payment that would be made when default occurs (this will use the unconditional default probabilities).

TABLE 1. Unconditional Default Probabilities

Time (Years)	Unconditional Default Probability	Survival Probability
1	0.0200	0.9800
2	0.0196	0.9604
3	0.0192	0.9412
4	0.0188	0.9224
5	0.0184	0.9039

Using s to represent the unknown swap spread that we need to ultimately solve for, the calculations are shown in Table 2 and Table 3.

TABLE 2. Protection Buyer's Payments — Annual Spread Payment

Time (Years)	Survival Probability	Expected Payment	Discount Factor	PV of Expected Payment
1	0.9800	$0.9800s$	0.9512	$0.9322s$
2	0.9604	$0.9604s$	0.9048	$0.8690s$
3	0.9412	$0.9412s$	0.8607	$0.8101s$
4	0.9224	$0.9224s$	0.8187	$0.7552s$
5	0.9039	$0.9039s$	0.7788	$0.7040s$
Total				$4.0704s$

TABLE 3. Protection Buyer's Payments — Accrual Spread Payment

Time (Years)	Default Probability	Expected Payment	Discount Factor	PV of Expected Payment
0.5	0.0200	0.0100s	0.9753	0.0098s
1.5	0.0196	0.0098s	0.9277	0.0091s
2.5	0.0192	0.0096s	0.8825	0.0085s
3.5	0.0188	0.0094s	0.8395	0.0079s
4.5	0.0184	0.0092s	0.7985	0.0074s
Total				0.0426s

From these two tables, the total expected swap spread payments equal to:

$$\text{Total Expected Swap Spread Payments} = 4.0704s + .0426s = 4.1130s$$

- **Step 3: Determine Protection Seller's Payments**

Finally, Table 4 on the following page shows the expected CDS payments from the seller's side. Here, we multiply the probability of default by one minus the recovery rate to get the expected payment, or $.02 * (1 - .4) = .0120$. We then multiply by the discount factor to get the discounted expected payment and add up for all possible default dates.

TABLE 4. Protection Seller's Payments

Time (Years)	Default Probability	Recovery Rate	Expected Payment	Discount Factor	PV of Expected Payoff
0.5	0.0200	0.4	0.0120	0.9753	0.0117
1.5	0.0196	0.4	0.0118	0.9277	0.0109
2.5	0.0192	0.4	0.0115	0.8825	0.0102
3.5	0.0188	0.4	0.0113	0.8395	0.0095
4.5	0.0184	0.4	0.0111	0.7985	0.0088
Total					0.0511

- **Step 4: Solve for the CDS Spread, s**

  To solve for the CDS spread, we set the buyer and seller side payments equal to each other and solve for s:

$$4.1130s = .0511 \Rightarrow s = 0.0124$$

Additional CDS Details

- Marking to Market — The calculations above were used to determine the CDS spread at inception, assuming an initial value of zero. The same calculations could be performed at any future date using the known CDS spread in order to mark an existing CDS to market.

- Default Probabilities — The default probabilities used are risk neutral default probabilities, as derived and explained in Hull Chapter 23. Alternatively, we could use actual CDS spreads to infer the implied risk neutral default probabilities. If we wanted to assume a constant conditional default probability (as we did above), then the math would be easy. More realistically, we would use spreads on different maturity CDS to "bootstrap" the default intensities for each year.

- Binary CDS — This is essentially the same, but the payoff is not a function of the recovery rate. Instead, the payoff is a fixed dollar amount.

- Recovery Rate — Notice that as long as the same recovery rate is used to calculate the risk neutral default probabilities and to value the CDS, the assumed recovery rate doesn't have a significant impact on the estimated CDS spread or the value of an existing CDS.

- CDS Market — The CDS market has grown tremendously in recent years, though its future growth is debated because of the potential for some parties to have more information about a particular company's default probability than other parties.

  It is important to note that, as we will address below, CDS contracts played a fairly critical role in fueling the credit and securitization bubble that burst so badly in 2007-2009. Since the universe of corporate bonds and even securitized tranches is relatively small, it was quite common for CDS contracts to be used in place of actual bonds in CDO structures. The ability to create hundreds or thousands of CDOs from the same underlying risk using CDS instead of the actual bonds partly contributed to the magnitude of the crisis. When a single corporate or structured bond defaulted, it tended to effect numerous (synthetic) CDO structures that didn't contain the actual bond but contained instead the CDS on that bond.

  Despite this role, and the bad press that came with it, the CDS market for single-name corporate credits actually functioned well during the crisis and proved to be a critical mechanism for firms to manage their credit exposure through the crisis.

Credit Indices

To track the overall CDS market, several indices have been created to show not only the average CDS spread for a portfolio of corporate credits but also to provide a means to trade (buy or sell protection) on entire baskets. Two examples of such indices include:

- CDX NA IG — This is a portfolio of 125 investment grade North American companies.
- iTraxx Europe — This is a portfolio of 125 investment grade names in Europe.

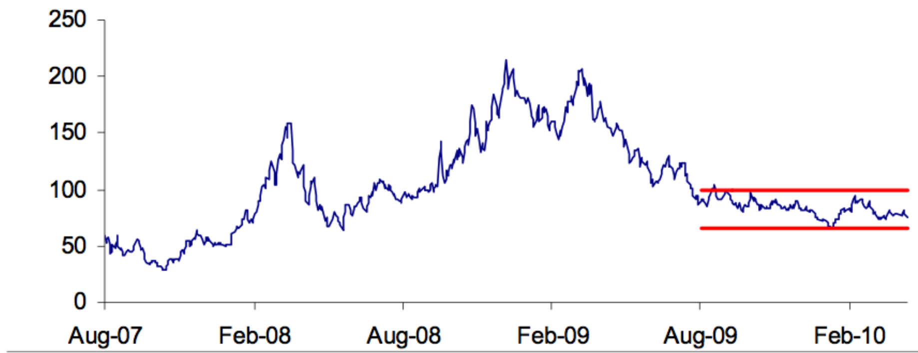
These indices are updated twice per year in then event a company in the index is no longer investment grade. Of course, these are just two examples. Numerous other indices capturing different sectors, geographies or credit ratings also exist.

Since the index spread approximately equals the average CDS spread for the companies in the index, it serves as a useful indicator of the market's perceived corporate default risk. For instance, consider the historical data for the iTraxx investment grade index from 2007 to 2010 shown in Figure 1 on the next page.

The important thing to note though is that the index can actually be traded, with participants either buying or selling basket protection on all 125 names in the index simultaneously. Further, specific tranches of these indices can also be traded.

Finally, note that the text briefly discusses some technical details of how CDS index transactions are quoted and traded. See Section 24.4 for further details.

FIGURE 1. iTraxx Europe 5-Year Spreads (in basis points)



Source: Credit Suisse

Other Credit Derivatives

CDS Forwards and Options

Once the CDS market became established, traders began to buy and sell CDS forward contracts and CDS option contracts. These give the parties either the obligation or the right, respectively, to enter into a CDS at a future date.

Basket Credit Default Swaps

These have a number of reference entities and provide a payoff either when any of them default, or when only the first one defaults, or only the second defaults, etc.

Total Return Swaps

These are agreements to exchange the total return on a bond, including coupons, interest and capital gains/losses for LIBOR plus a spread. One party, the swap receiver, pays LIBOR plus a spread every six months and receives the coupons from the reference bond every six months. On the final swap date, the capital gain/loss on the bond is paid from the payer to the receiver. If there is a loss, the payment is made from the receiver to the payer.

The main reason for doing this type of swap is so that one party can "invest" in a risky bond without having to actually pay for it up front. Instead, they have another party, the swap payer, buy the bond and the swap receiver agrees to make LIBOR plus a spread payments to the payer much as they would do if they had borrowed the money from the payer in the first place to buy the bond on their own. The payer's swap payments then pass along the economic interest in the bond to the receiver, so that if the bond suffers a default or other loss in value, the receiver suffers.

Because this is much like an exchange of a LIBOR bond for the reference bond, we might expect the spreads to be set to equate the values of these two bonds. However, in the event of the default of the receiver, the payer will lose money if the reference bond had declined in value and therefore the payer will demand a spread to LIBOR based on the receiver's credit quality and the correlation of default between the receiver and the reference bond.

Collateralized Debt Obligations (CDOs)

The text begins with a discussion of asset-backed securities and CDOs, which we have sufficiently covered in other readings. But focus on a few points not made in those readings:

Cash CDO vs. Synthetic CDO

In the other readings, we typically suggested that the underlying instruments in the pool of *debt obligations* were either corporate bonds or tranches (typically the hard-to-sell mezzanine tranches) from other asset-backed securities. When the pool consists of actual bonds like this, it is referred to as a **Cash CDO**. Because actual bonds are held, typically the CDO will not be entirely static. Subject to a variety of criteria, often the CDO manager will have the ability to sell bonds within the pool and replace them with other bonds.

But recall that the objective of a CDO is largely to transfer credit risk. As a result, there is really no need for the pool to contain actual bonds and instead could consist solely of credit default swaps. In such a case, we refer to the resulting structure as a **Synthetic CDO**. To construct this, as before the various tranche investors purchase bonds from the special purpose vehicle (SPV) and expect to get both principal and interest back over time. But rather than use the proceeds from these bonds sales to buy corporate bonds or ABS tranches for the pool, the SPV *sells* credit protection to buyers via short positions in CDS contracts. The CDS spreads paid are passed through, in some fashion in accordance to the relative risk they assumed, to the investors in the various CDO tranches. The proceeds from the CDO investors (the ultimate credit risk takers) serve as collateral for the CDS contracts and if a default occurs, payments are made out of these funds.

Just as in a cash CDO, the investors suffer losses sequentially, with the equity tranche investors losing their principal when the first $\$x$ of defaults occur, the mezzanine tranche investors losing their principal when the next $\$y$ defaults occur, etc.

Single Tranche Trading

As mentioned earlier, the CDX NA IG and iTraxx indices track the CDS spreads on a pool of CDS contracts. When investors look to buy or sell protection on a specific tranche (layer) of this index, this trade is referred to as *single tranche trading*. These are standardized trades, with predefined tranche sizes/layers.

The Role of Correlation

As mentioned extensively in the Coval, *et. al.* reading, the expected cash flows for a CDO tranche depend critically on the correlation across the underlying debt instruments in the pool. When the underlying instruments in a large pool of debt obligations are independent, the expected defaults on the junior tranches are very high and the expected defaults on the senior-most tranches are low. But as the correlation increases, then each of the tranches become more like each other. In the limit with perfect correlation (and no recovery), every tranche must have the same default rates as the underlying assets. Therefore, the expected defaults for the junior tranches decline and the expected defaults for the senior tranches rise as the correlation increases.

Practice Questions

Question 1. Determine the spread for a 2-year CDS on a bond with a face value of $\$1,000$. Assume that the reference bond's conditional default probability is 3% per annum, that default always occurs at the *end* of each annual period and that the spread is also paid at the end of each period. In the event of default there is a 30% recovery rate and the accrued spread payments are paid upon default. The risk free rate is 4%, continuously compounded.

Solution. Note that this is nearly identical to the problem in the text and in the notes above. In addition to changing the assumptions though, I have also made the default occur at the end of the period to slightly reduce the number of calculations. Otherwise the steps are the same, as shown here.

- Step 1: Calculate the Default Probabilities

  First, using the conditional default probability of 3% per year, we determine the unconditional default probabilities and the survival probabilities. For the first year, this is easy — 3% default probability and 97% survival probability. For the second year, the unconditional default probability is $.97 * .03 = .0291$ and the survival probability is $.97 * .97 = .9409$.

TABLE 5. Unconditional Default Probabilities

Time (Years)	Unconditional Default Probability	Survival Probability
1	0.0300	0.9700
2	0.0291	0.9409

- Step 2: Determine Protection Buyer's Payments

  Next, calculate the expected CDS spread payments. The spread of s is paid each period assuming that there is survival and then in the event of default the accrued spread payment is also made. Here, because I assumed default occurs at the end of the year, the accrued payment is also s.

TABLE 6. Swap Spread Payments — Annual

Time (Years)	Survival Probability	Expected Payment	Discount Factor	PV of Expected Payment
1	0.9700	0.9700s	0.9608	0.9320s
2	0.9409	0.9409s	0.9231	0.8686s
Total				1.8005s

TABLE 7. Swap Spread Payments — Accrual

Time (Years)	Default Probability	Expected Payment	Discount Factor	PV of Expected Payment
1	0.0300	0.0300s	0.9608	0.0288s
2	0.0291	0.0291s	0.9231	0.0269s
Total				0.0557s

The total expected spread payments are therefore:

$$1.8005s + .0557s = 1.8562s$$

- Step 3: Determine the Protection Seller's CDS Payments

  The expected CDS payoffs are calculated similarly. Here, we multiply the probability of default by one minus the recovery rate to get the expected payment, or $.03 * (1 - .3) = .0210$. We then multiply by the discount factor to get the discounted expected payment and add up for all possible default dates.

TABLE 8. CDS Payments

Time (Years)	Default Probability	Recovery Rate	Expected Payment	Discount Factor	PV of Expected Payoff
1	0.0300	0.3000	0.0210	0.9608	0.0202
2	0.0291	0.3000	0.0204	0.9231	0.0188
Total					0.0390

- Step 4: Solve for s

Set the two sides equal to each other and solve for s:

$$1.8562s = .0390 \Rightarrow s = .0210$$

That is, for each dollar of notional value the CDS spread is $s = 210$ basis points.

Question 2. Assume the same facts as in Question 1, except that default always occurs midway through the year at $T = .5$ or $T = 1.5$. What would the CDS spread be in this case?

Solution. This is the way Hull did the questions in the textbook. It differs from my answer above because of the size of the accrued spread payment (only half the spread is owed) and the discount factors used for both the accrual payment and the protection seller's payments.

The full solutions is as follows:

Step 1: Calculate the Default Probabilities

First, using the conditional default probability of 3% per year, we determine the unconditional default probabilities and the survival probabilities. For the first year, this is easy — 3% default probability and 97% survival probability. For the second year, the unconditional default probability is $.97 * .03 = .0291$ and the survival probability is $.97 * .97 = .9409$.

TABLE 9. Unconditional Default Probabilities

Time (Years)	Unconditional Default Probability	Survival Probability
1	0.0300	0.9700
2	0.0291	0.9409

Step 2: Determine Protection Buyer's Payments

Next, calculate the expected CDS spread payments. The spread of s is paid each period assuming that there is survival and then in the event of default the accrued spread payment is also made. Here, because I assumed default occurs at the middle of the year, the accrued payment is $.5s$.

TABLE 10. Swap Spread Payments — Annual

Time (Years)	Survival Probability	Expected Payment	Discount Factor	PV of Expected Payment
1	0.9700	$0.9700s$	0.9608	$0.9320s$
2	0.9409	$0.9409s$	0.9231	$0.8686s$
Total				$1.8005s$

TABLE 11. Swap Spread Payments — Accrual

Time (Years)	Default Probability	Expected Payment	Discount Factor	PV of Expected Payment
.5	0.0300	$0.0300(.5s)$	0.9802	$0.0147s$
1.5	0.0291	$0.0291(.5s)$	0.9418	$0.0137s$
Total				$0.0284s$

The total expected spread payments are therefore:

$$1.8005s + .0284s = 1.8289s$$

Step 3: Determine the Protection Seller's CDS Payments

The expected CDS payoffs are calculated similarly. Here, we multiply the probability of default by one minus the recovery rate to get the expected payment, or $.03 * (1 - .3) = .0210$. We then multiply by the discount factor to get the discounted expected payment and add up for all possible default dates.

TABLE 12. CDS Payments

Time (Years)	Default Probability	Recovery Rate	Expected Payment	Discount Factor	PV of Expected Payoff
.5	0.0300	0.3000	0.0210	0.9802	0.0206
1.5	0.0291	0.3000	0.0204	0.9418	0.0192
Total					0.0398

Step 4: Solve for s

Set the two sides equal to each other and solve for s:

$$1.8289s = .0398 \Rightarrow s = .0217$$

That is, for each dollar of notional value the CDS spread is $s = 217$ basis points.

Question 3. Assume that for the CDS in Question 1, default actually occurred at time $t = .75$ and the recovery rate was actually 35% of the face value. What payments would be made by the buyer and the seller of the swap on that date?

Solution. For the seller of the swap, they pay the difference between par and the recovery rate times the face value, or $\$1000 - 35\% * \$1000 = \$650$.

For the buyer, their payments occur annually in this question, however if default occurs in between those dates they owe the seller of protection their accrued payment. Therefore, they pay the accrued swap payment, $.75 * 2.1\% * 1000 = 15.75$.

Question 4. Hull notes that a critical input to the determination of the expected cash flows in a CDS is the recovery rate estimate that we cannot observe. However, for plain vanilla credit default swaps the CDS spread is generally insensitive to different assumptions about the recovery rate. Why is this the case?

Solution. Because the recovery rate assumption enters into the pricing of the CDS in two largely offsetting ways. It affects the risk neutral default probabilities and it affects the payoffs in the event of default.

It affects the risk neutral default probabilities because in practice we use the prices of bonds and an assumed recovery rate to infer the default probability. The price of the bond tells us what the market thinks the expected present value of default losses will be. If we overstate the recovery rate in this exercise, we will understate the loss given default and since the expected present value of default losses is fixed, we will overstate the default probability.

Then, when we use these same overstated recovery rates and overstated default probabilities to determine the CDS spread, the expected payoffs for the CDS will be largely unaffected.

Question 5. Suppose a basket of 100 corporate bonds is created, each of which has a 2% probability of defaulting over the next 5 years. We are going to use this basket to create a credit derivative that pays \$100 million if there is one or more defaults within the next 5 years. If the defaults are uncorrelated across these 100 reference entities, what is the expected payoff for the protection buyer?

Solution. We simply need to know the probability that one or more of these bonds defaults. Since the defaults are uncorrelated we can use the binomial distribution to get the following expected payoff:

$$\begin{aligned}
\text{Expected Payoff} &= \$100 * \text{Prob}(1 \text{ or more defaults}) \\
&= \$100 * [1 - \text{Prob(zero defaults)}] \\
&= \$100 * \left[1 - \binom{100}{0}(.02)^0(.98)^{100}\right] \\
&= \$100 * (86.74\%) \\
&= \$86.74 \text{ million}
\end{aligned}$$

Question 6. Suppose we create a CDO containing 100 corporate bonds with 5-year maturities, each of which has a 2% probability of defaulting over the next 5 years, a notional value of \$1 million and a recovery rate in the event of default of 0%. If the bond defaults are perfectly correlated ($\rho = 1.0$), what is the expected default loss in dollars for investors in the super-senior tranche that covers losses from 60% - 100% over the next 5-years?

Solution. This is similar to the previous question, but now with the defaults perfectly correlated all tranches experience the same default rates because either none of the bonds defaults, in which case the default losses are zero for all tranches, or all of the bonds default, in which case the default losses are 100% for all tranches. Since there is a 2% probability of default, the expected losses in the super-senior tranche are also 2% of the notional. Here, the total CDO notional is $100 million and the super-senior tranche represents the top $40 million of the CDO structure. Therefore, the expected default loss in dollars is $0.8 million.

Question 7. Using the same information as in the previous question, but assuming that the recovery rate in the event of default is 40% for all bonds in the pool (i.e. the loss in the event of default is 60% of the face value), what are the expected default losses in dollars for the super-senior tranche (covering the 60% - 100% layer of loss)?

Solution. In this case, whenever a bond defaults the pool still recovers 40% of the notional value. As a result, even if EVERY bond were to default the pool would still collect, over the five years, $40 million. As a result, there is no way for the super-senior tranche to lose a single dollar. The expected default losses are zero.

Notice something important here. While it is true that understating the correlation of bonds in a CDO could result in substantial understatement of the probability of default of the super-senior tranches, unless recovery rates are zero there could indeed be a layer high enough that is perfectly insulated against loss.

Question 8. During the credit crisis AIG's Financial Products subsidiary faced massive mark-to-market losses (and collateral calls) on their super-senior exposure to ABS CDOs that contained mezzanine (BBB-rated) tranches of subprime mortgage backed securities. They argued at the time that it was inconceivable that they could lose money on their exposure in part because, like in the previous question, they were very senior in the credit structure and so could only lose money if substantially all of the underlying bonds defaulted *and* the recovery rates were extremely low, which they believed to be impossible.

What characteristics are unique to CDOs of mezzanine tranches of subprime securitizations that make the probability of these two conditions much higher than AIGFP argued was the case?

Hint: Mezzanine tranches of subprime MBS are rather difficult to sell and so great care is taken to keep those tranches as small as possible. Often the BBB-rated tranches might cover a loss range of only 1%.

Solution. If AIGFP had been talking about their CDOs of corporate bonds, and if their attachment points were high enough, then much like in the previous question they would perhaps correctly thought that recovery rates, even in the event of default, would be relatively high.

However, as noted in the hint above, the BBB-rated tranches of subprime MBS were very *thin* so that recovery was more likely to be completely binary - either it has no default or complete default with no recovery. This particular feature made it inappropriate to assume recovery rates similar to historical recovery rates on corporate bonds, which are in the 40% - 60% range.

In addition, as discussed in the Coval reading, all subprime mortgages were highly dependent upon home price appreciation in order to avoid the almost certain defaults that would occur if the homeowners actually had to continue making their payments. This caused all subprime MBS mezzanine tranches to be highly correlated.

The combination of high correlation across the instruments in the pool and the low recovery rates in the event of default (due solely to the thinness of those tranches) make this situation very different from the one in the previous question. There, it was impossible for the top 40% of the structure to experience a loss; here it was conceivable that every mezzanine tranche could default with zero recovery, causing complete default for even the senior-most tranche of the ABS CDO.

Recommended Textbook Problems

I strongly recommend working all of the end-of-chapter questions from the Hull textbook, except for those that relate to Section 24.10 and Section 24.11 which is not on the syllabus. But due to time constraints this may not be feasible. Therefore, at a minimum you should review all of the numerical examples that appear throughout the main text and the following end-of-chapter questions (especially the ones in bold):

2, 5, 6, **8, 9**, 10, 11, 14

Cummins: CAT Bonds and Other Risk-Linked Securities

Early Developments in Risk-Linked Securities

After Hurricane Andrew in 1992, there were numerous attempts to change the insurance industry's reliance on the reinsurance market as the sole vehicle for managing property catastrophe risk. Some notable examples included various attempts to develop exchange-traded futures and options on catastrophe indices and contingent surplus notes.

CBOT Futures and Options and BCE Options

Both the Chicago Board of Trade and the Bermuda Commodities Exchange attempted to develop futures and options contracts tied to industry loss indices (using PCS loss indices). These contracts never were able to gain much traction for several reasons, including limited number of market participants, questions of counterparty credit risk, concerns about disrupting reinsurer relationships and, most notably, concern regarding the degree of basis risk between company-specific and industry index losses.

Contingent Surplus Notes

Another early attempt was the contingent surplus note, first issued by Nationwide and then by two other insurers. In this transaction, a trust was established to issue 10-year securities paying a coupon of 220 basis points above US Treasury rates and invest the proceeds into US Treasury bonds. However, Nationwide retained the option to substitute surplus notes (bonds issued by insurers which state insurance departments allow to be counted as surplus because they are rank lower in priority than policyholder obligations) paying a coupon of 9.22% if an event (such as, but not limited to, a property catastrophe event) occurred. So, in the event of a large loss of surplus, Nationwide was guaranteed the ability to raise $400 million of cash through the issuance of surplus notes.

While interesting, this approach was not very appealing to investors since it was not a clean bet on the occurrence of a property catastrophe event and they ultimately faced the credit risk of Nationwide to repay the surplus notes. For Nationwide, this provided assurance of post-event financing, but unlike reinsurance, it carried with it an obligation for the funds to be repaid.

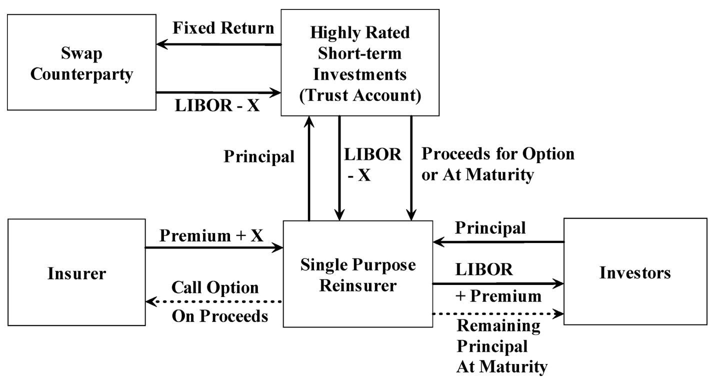
Catastrophe Bonds

This is the area that has received the greatest attention and has had the most success to date. New issuances peaked in 2007, shortly after Hurricane Katrina, and then fell noticeably during the financial crisis. It has since picked up again, with issuance in 2014 totaling more than $8 billion and exceeding the previous 2007 peak.

Basic Structure of a Catastrophe Bond

Catastrophe bonds act exactly like catastrophe reinsurance, but use a specific structure that allows the risk to be securitized (turned into traceable bonds rather than private catastrophe reinsurance contracts) and assumed by investors such as pension funds or hedge funds. The basic structure is much like a mortgage backed security (discussed in the BKM readings), as depicted in Figure 1, taken from the Cummins paper.

FIGURE 1. Basic Catastrophe Bond Structure



Description of Key Elements of Catastrophe Bond Structure

Focusing for the moment only on the bottom three boxes in the diagram, there are two critical aspects of a catastrophe bond.

First, as shown through the two boxes on the right, investors provide capital (to purchase a bond) to a Single Purpose Reinsurer (SPR). They pay for the bond by paying the par value (the "principal" amount shown in the diagram) and in return they get a promise to repay the principal with interest at the maturity of the bond. The SPR will repay the principal at maturity only if they have sufficient assets to make the payment. Otherwise, they will default, in full or in part, on their promise to repay the principal at maturity.

Second, and simultaneous with the above transaction, an insurer purchases catastrophe reinsurance from the SPR. The form of that reinsurance contract will look essentially identical to a traditional reinsurance contract. The SPR will pay claims covered by the contract, which is the main reason why the SPR may not have sufficient funds at the maturity of the bond to repay investors.

As the result of those two parts of the catastrophe bond structure, the catastrophe risk faced by the insurer is transferred to investors, who ultimately bear the risk of loss from a catastrophe event. Remember, the SPR is holding onto collateral, so they can repay the principal at maturity unless they have to use some or all of it to pay a covered claim under the reinsurance agreement. Since the reinsurer is set up with a *single purpose*, it is essentially barred from any other activities (see the swap discussion below) and so the occurrence of a catastrophe event that causes a covered claim is the only event that can cause a default on the bonds.

Because the SPR is a licensed reinsurer, the insurer treats this transaction as it would any other reinsurance contract. But unlike traditional reinsurance, the insurer obtaining the reinsurance does not have to be concerned with the credit risk of the reinsurer because the full limit of the reinsurance policy is *collateralized* by the proceeds from the bonds.

Notice in the diagram that the insurer pays a reinsurance premium to the SPR, just as they would pay a traditional reinsurer. They also pay an additional amount, denoted ”X” in the diagram, which could be used to cover fees associated with setting up SPR but is also used to make up any difference between the floating risk free rate of interest (LIBOR) that the investors want to earn on the collateral they provided. The reinsurance premium provides the investors with the compensation they require to assume the catastrophe risk and the additional amount, ”X”, ensures that they also earn LIBOR on the funds.

Investment Activities of the Special Purpose Reinsurer

The top part of the catastrophe bond diagram addresses interest income that is earned on the collateral while it is being invested.

As suggested just above, the investors are providing collateral up-front and expect to earn interest income on that collateral. One approach might be to just invest the money in very short-term, highly rated and floating rate assets so that the investors are not facing any risk of loss of principal, are not facing interest risk and are not facing credit risk on the invested assets. The only risk they would face is the risk of a covered catastrophe event reducing the collateral available to pay them principal at maturity.

In fact, that is exactly what is now done. Most catastrophe bonds issued today allow the SPR to only invest in money market funds. But prior to the financial crisis it was common for SPRs to invest in riskier assets - fixed coupons, longer maturity and lower credit rating. Knowing that this riskier investment strategy could lead to loss of principal from events other than catastrophe events, the investors would also require the SPR to enter into a Total Return Swap with a highly rated swap counterparty to make-up any losses on the invested collateral.

This *total return swap* accomplished two goals:

i. swap the fixed interest receipts from its investments into floating rate interest receipts

ii. provide protection against a loss in market value of the collateral in the event it is invested in assets that are not completely free of credit risk or that contain interest rate risk.

This investment of the collateral into “Highly Rated Short-term Investments” and the simultaneous use of a total return swap is what is depicted in the top two boxes in the diagram.

Although not addressed in the paper, it is important to note that the bankruptcy of Lehman Brothers in 2008 caused substantial upheaval in the catastrophe bond market. The issue, in the case of four specific bonds, was that Lehman Brothers was the counterparty in the total return swap and responsible for guaranteeing the periodic LIBOR interest as well as protecting the value of the underlying collateral assets. Had the assets been invested in relatively risk-free assets, the Lehman bankruptcy would have been addressed by replacing them as the total return swap counterparty. However, Lehman had been stuffing the collateral account with toxic assets that were worth much less than their par value at the time that Lehman filed for bankruptcy. Neither the sponsors nor the investors seemed to have been paying much attention to the assets held in the collateral account, as they both assumed that Lehman would be responsible for any resulting losses in the value of the collateral and/or the periodic interest earned. But Lehman's bankruptcy meant that the insurers purchasing protection, in the event they had a covered claim, and investors were forced to reassume the market and credit risk on the collateral, which just happened to be worth substantially less than par at the time of the bankruptcy.

Since that time, total return swaps have generally not been used. As noted above, collateral is now more commonly held in money-market funds with lower returns and much lower risk.

Catastrophe Bonds vs. Traditional Reinsurance

A few things are important to note about catastrophe bonds relative to traditional reinsurance.

- Collateral — The policy limit is collateralized and so the insured does not face the credit risk of the reinsurer, unless the collateral is lost as the result of poor investments. This risk is easy to control by just prohibiting such investments.

- Multi-Year Coverage — Catastrophe bonds are typically issued for multiple risk periods (3 year periods). This locks in pricing and also serves to amortize the significant fixed expenses associated with issuing the bonds (due to banking fees, legal costs, rating agency costs, risk modeling costs, etc. which can amount to more than $2 million up-front and $250,000 per annum).

- Risk-remote Layers — Catastrophe bonds are typically issued for more risk remote layers (roughly in the range of 1% - 5% probability of occurrence).

- Indemnity vs. Index Basis — Some catastrophe bonds cover insured losses on an indemnity basis, which reimburses the sponsor for actual losses incurred. However, the majority of catastrophe bonds issued to date have actually provided coverage on an index basis, with payouts determined based on aggregate industry losses, modeled losses or parametric factors.

- Pricing — Because the investors are pension funds and hedge funds without any natural exposure to insurance risks, it is *arguably* possible for the pricing to reflect a lower risk margin over and above the expected loss than in the traditional reinsurance market (in practice, this is not usually the case). In addition, investors see these as a pure play bet on catastrophe risk, since the bonds are not issued directly by the insurer and are segregated from the rest of its business. They only face the catastrophe risk and not the general business risk of the insurer.

Loss Triggers

There are several ways to design the *loss triggers* that determine the circumstances and amount of the covered claims under the reinsurance treaty.

Indemnity Trigger

This works just like in traditional reinsurance, covering the actual losses of the insured, subject to the specified retentions and limits per occurrence or in the aggregate. This does not result in basis risk for the insured, but it does complicate the risk modeling for the investors, requiring them to more carefully underwrite the risks and understand the nuances of the reinsurance business.

Index Trigger

An index trigger is intended to pay claims based upon an index that is not tied directly to the insured's actual losses. Three specific forms are used.

- *Industry index loss triggers* rely on aggregate industry losses from specified events (e.g. hurricane, earthquake) in specified geographies. Within the U.S. the Property Claims Service (part of ISO) indices are commonly used, but other customized indices can be used as well, including specific weightings of various PCS indices.

- *Modeled loss triggers* apply an agreed upon catastrophe risk model, using the physical parameters of the actual events, to the insured's exposure data and pay according to the model's estimated losses rather than actual losses.

- *Parametric triggers* pay fixed dollar amounts based upon the physical characteristics of catastrophe events, such as wind speed for hurricanes or magnitude of earthquake.

Hybrid Trigger

Hybrid triggers simply blend multiple triggers into a single bond.

Choice of Loss Triggers

The choice of loss triggers is a matter of balancing the trade-off between moral hazard, transparency for the investors (which ultimately impacts investor demand, ease of execution, etc.) and basis risk for the insured. The following summarizes the key considerations along these dimensions:

Indemnity Trigger

Indemnity triggers generally eliminate the basis risk for the insured, but require much more extensive disclosure to investors, require much more investor due diligence and complicate claim settlement after an occurrence. The latter issue can be important because of the need to ensure that the SPR is on automatic pilot with a contractual maturity date of the bond.

Index Triggers

Index triggers are easier for investors to evaluate because they are not subject to the claims handling practices of the insured and therefore the risk analysis is simplified. Index-based bonds are also not impacted by changes in risk exposure, premium volume, etc., so the overall due diligence process is simplified and execution risk is minimized. However, this comes with increased basis risk, since it is possible that the index losses do not mirror the company's actual claim experience.

Sidecars

Sidecars are quite similar to catastrophe bonds in the sense that they also use a special purpose reinsurer to issue a reinsurance policy funded by capital markets investors. But unlike catastrophe bonds, which are structured as excess of loss contracts, sidecars are structured to provide quota share reinsurance capacity, either for a reinsurer's entire book of business or some subset thereof.

A key difference between sidecars and catastrophe bonds is the fact that the financing for sidecars typically includes both equity and debt tranches. The debt tranches are fairly indistinguishable from catastrophe bonds, but the equity tranche investors retain ownership of the residual assets of the SPR once the bonds are repaid and therefore have the potential for more upside returns.

From the issuer or insured's perspective, a sidecar is used as a means to obtain temporary quota share capacity. For this reason, significant amounts of sidecar capital was raised after hurricanes Katrina, Rita and Wilma in 2005.

Although not addressed in the paper, you should note that the term sidecar is used to refer to transactions like those described above in which the insured is ultimately seeking quota share capacity as well as transactions that are essentially collateralized debt obligations that pool reinsurance treaties as collateral for various securitized debt and equity tranches (these are increasingly being referred to as collateralized reinsurance obligations).

Catastrophe Equity Puts (Cat-E-Puts)

These products were private transactions between insurers and reinsurers to provide contingent financing after a catastrophe event. If the defined event occurred, the insurer was able to raise additional capital from the reinsurer on pre-agreed terms, essentially a put on its own equity, in the form of preferred stock.

Cat-E-Puts are a specific form of a contingent capital transaction that uses preferred stock as the instrument. Similar transactions in which debt, surplus notes or even common equity can also be structured. The key point is that there is a defined event that needs to occur before the financing is provided, but the terms of that financing are negotiated before the event occurs.

Unlike a catastrophe bond or traditional reinsurance, which pays the insured for their covered claims with no contractual *payback*, a contingent capital transaction provides the same funds to cover claims but results in a contractual obligation to repay investors principal and interest over time. This is really just *post-event financing* of the claims as opposed to *risk transfer*.

Catastrophe Risk Swaps

Note that there is no economically meaningful distinction between a reinsurance contract and a swap contract. In both cases the parties exchange fixed and floating payments over time. In the case of reinsurance, the fixed payments are represented by the premiums and the floating payments are represented by the covered claims. It is therefore possible for two parties, such as an insurance company and a hedge fund, to enter into what would otherwise look like a reinsurance contract but use the legal form of a swap contract. The cash flows would be the same, but the regulatory and accounting implications may differ.

Taking that idea one step further, it may be possible for two insurers to enter into a swap whereby they each swap floating rate payments tied to different indices — such as one party paying according to industry-wide or company-specific California earthquake claims and the other party paying according to industry-wide or company-specific Florida hurricane claims. If the expected losses are similar, they may even be able to do this with no net up-front premium, but in general they could also swap the up-front premium payments too. This type of transaction is often called a *catastrophe risk swap* or an *exposure swap*.

Industry Loss Warranties (ILW)

ILW's are essentially index-based reinsurance agreements, where the payoff is intended to be based on industry-wide claims from defined perils, in defined regions, and for defined classes of business (e.g. personal lines claims from hurricane losses in Florida). These appeal both to insurers looking to hedge their risk as well as investors who simply want to bet on the occurrence of catastrophic events. In the case of insurers hedging risks, they of course face basis risk as in an index-based catastrophe bond.

ILW's can be written in derivative or swap form. In many cases the party buying protection is an insurance company, in which case there are regulatory and accounting reasons for them to want the contractual form to resemble a reinsurance agreement. Among other things, this requires the payout to be limited to the ultimate net loss incurred by the insurer. For this reason, ILW's typically have *dual triggers*, with both an industry loss component (sometimes called the warranty trigger) and a company-specific ultimate net loss component (sometimes called the retention trigger). What isn't made clear in the paper though is the fact that often the ultimate net loss component is set at very low levels such that it is almost certain to be hit in the event the industry loss trigger is hit.

Payouts on an ILW can either be binary or proportional. In the case of the former, the full notional amount is paid if the trigger is hit. In the case of the latter, the notional amount is multiplied by the portion of losses within a specified layer of industry attachment and exhaustion points.

Between July 1, 2014 and June 30, 2015, ILW trading volume exceeded $4 billion, making the ILW a fairly significant source of risk capital.

Catastrophe Bond Market

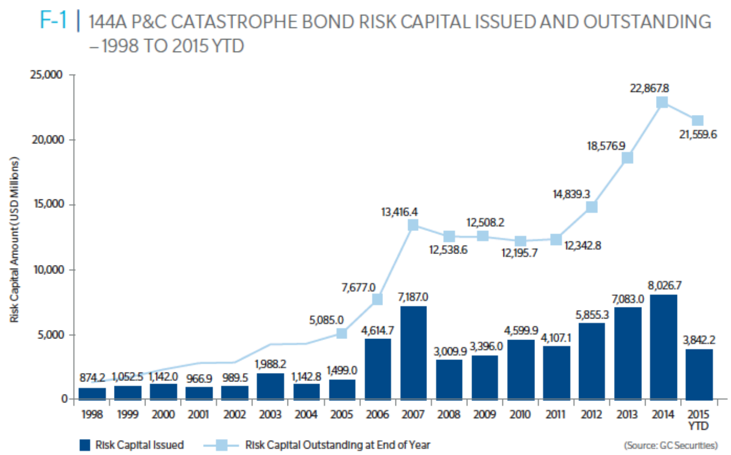
The paper presents several statistics regarding the size of the catastrophe bond market and characteristics of the bonds that have been issued to date, including typical ratings, expected default rates/expected default losses, perils, geographies, etc. Below are some highlights of that discussion, but note that this paper was written many years ago, so the statistics it presents may no longer be representative of the market. In a few places below I will provide more current market information.

Volume of New Issuance

Because a large catastrophe event would still represent a small fraction of the overall U.S. securities market, and because securities markets are generally more efficient than private markets (e.g. reinsurance), it has long been expected that the catastrophe bond market would be fairly large. In 2015, catastrophe bonds represent approximately 4% of global reinsurer capital. Including ILWs, sidecars and other forms of collateralized reinsurance, these forms of non-traditional property catastrophe capacity represent about 12% of global reinsurer capital.

Figure 2 on the next page shows new issuance volume through 2007, as presented in the paper, as well as some more recent data through Q2 2015.

FIGURE 2. P&C Catastrophe Bond Issuance:1998–2015



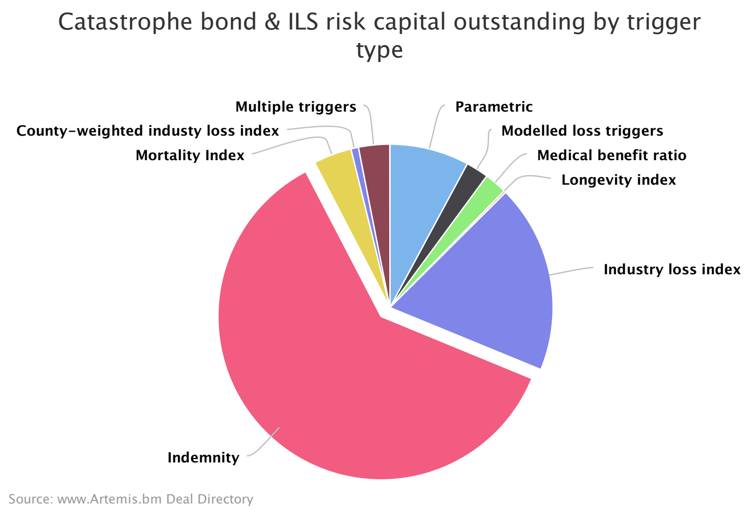
The credit crisis and other factors caused the volume of new issuance to fall significantly, but issuance has since rebounded and exceeds the 2007 level.

Note that when you factor in issuances of sidecar debt and equity, ILW's and other forms of collateralized reinsurance, the overall market for securitization of insurance risk is more significant than shown here.

Trigger Types

The trigger types have varied somewhat significantly from year to year, with index-based bonds tending to dominate the market in the early years due to their general appeal to investors. Indemnity bonds became more common from 2005 – 2008, particularly after many national primary carriers with mostly residential exposures were able to issue indemnity bonds. By 2015, indemnity bonds dominated new issuances (note that the chart below includes some life/health bonds as well as property catastrophe bonds):

FIGURE 3. ILS Securities by Trigger Type



Tenor

Most bonds are issued for 3-year risk periods. This allows the issuers to lock in rates (albeit with reduced aggregate coverage) and also allows the up-front expenses associated with the issuances to be amortized over multiple risk periods.

Issuers

To date, the issuance by primary insurers and reinsurers has been nearly equally split, with minimal issuance by non-insurer corporations and governments.

Ratings

It is important for catastrophe bonds to have credit ratings by S&P, Moody's or one of the other rating agencies. In each case, the rating agencies assign ratings roughly in line with aggregate corporate bond default rates — recall that when a claim is paid by the special purpose reinsurer the direct impact on investors is a default (full or partial) on the catastrophe bond. The vast majority (over 90%) of bonds are issued with below-investment grade ratings (BB or B), but recently there have been more issuances of investment grade (BBB) tranches.

Investor Type

In the early years, a significant portion (55%) of catastrophe bonds were actually purchased by insurers and reinsurers, which did not technically represent a *transfer* of insurance risk to the capital markets. But in 2015, dedicated catastrophe bond funds accounted for 47% of purchases, followed by pension funds and other institutional investors (32%), mutual funds (9%), reinsurers (10%) and hedge funds (2%).

Loss Experience

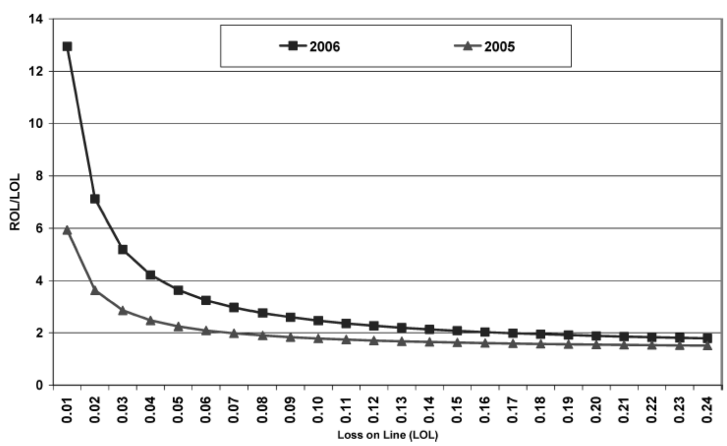
Investors in catastrophe bonds have tended to do well so far, with only one bond resulting in a total loss of principal. The Kamp Re bond issued in 2005 covered a significant amount of commercial risks in Louisiana and suffered a total loss from Hurricane Katrina. Rumors of inadequately modeled risks for Katrina (driven by the commercial exposure) initially caused some investor wariness of indemnity cat bonds, but interest in indemnity bonds (primarily covering residential exposures of primary insurers) rebounded and the longer term impact of Kamp Re is arguably positive, with the market now being "tested".

Catastrophe Bond Pricing

Since investors contribute funds to collateralize the policy limit, with these funds invested to earn short-term, risk-free returns, the net compensation for assuming the insurance risk is reflected in the coupon's spread over the risk-free reference rate. This spread is identical to the reinsurance premium normally paid for traditional reinsurance.

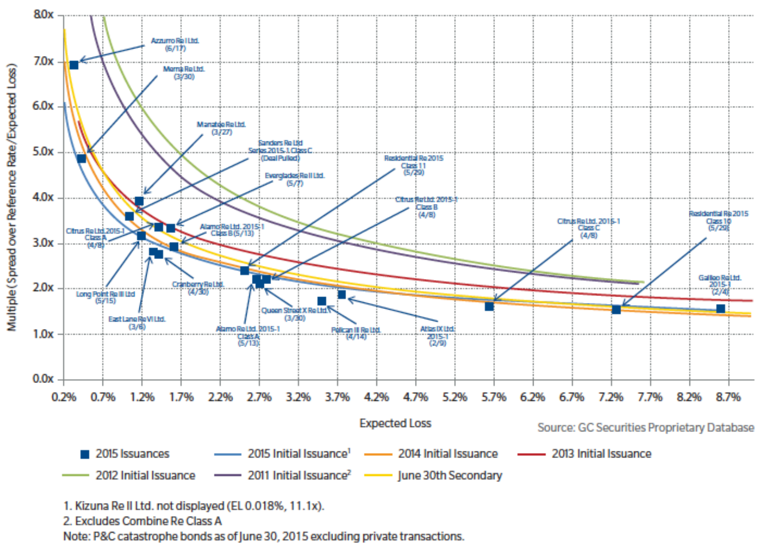
In the reading, the spread is compared to the expected losses on the bond (both relative to the bond limit, so shown as rate-on-line vs. expected loss-on-line). At the time, bonds with relatively low expected losses of 2% of the limit had spreads that were in the range of 4-7 times as large (8-14%), while bonds with expected losses in the range of 10% of the limit had spreads approximately 2 times as large (20%).

FIGURE 4. Cat Bond Pricing 2005-2006



Below is an updated picture of the cat bond spreads compared to the expected losses:

FIGURE 5. Cat Bond Pricing (2015)



Catastrophe Bonds vs. Traditional Reinsurance

As in any risk transfer transaction, the pricing will reflect the expected claims and expenses for investors, as well as a risk margin. In theory, this risk margin should be a function of the potential deviation from the expected claims and should be higher for investors with existing portfolios that have exposure to similar risks. For this reason, it is theoretically possible for non-insurer investors with, for instance, no natural exposure to Florida hurricane risk to price in a lower risk margin for a Florida hurricane catastrophe bond than would an existing reinsurer.

So, are cat bonds a cheaper form of risk transfer than traditional reinsurance? Not really, for two reasons.

- *Expenses* — The expenses associated with issuing a catastrophe bond are generally higher than placing reinsurance.

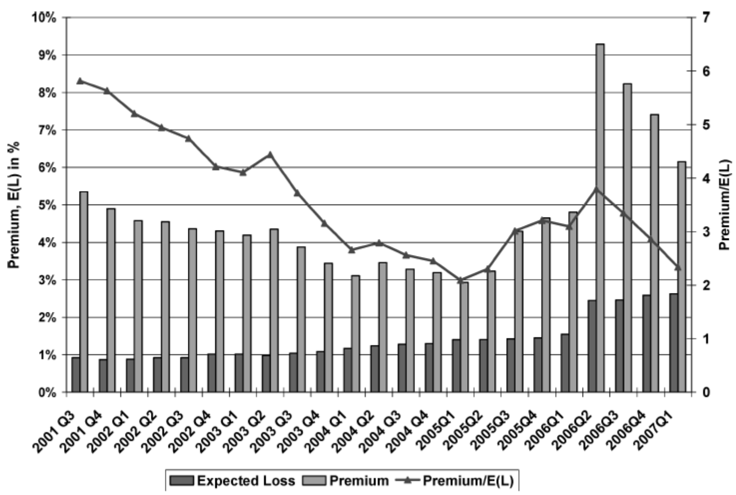
- *Value Pricing* — More importantly, the investors are providing a valuable risk transfer service to the issuers and therefore it is appropriate for the investors to charge a premium/spread that reflects the risk transfer value to the issuer and not simply the investor's minimum acceptable risk margin. That is to say, while investors arguably could rationally charge less, there's no real reason to do so. As long as cat bond pricing remains competitive with the traditional reinsurance market, there's no need for spreads to be reduced further.

Compared against traditional reinsurance market pricing, catastrophe bond pricing has become very competitive. Coupled with the overall decline in transaction costs and execution uncertainty, catastrophe bonds are now a viable alternative to traditional reinsurance for many insurers.

Historical Catastrophe Bond Pricing

The paper examines the spreads on catastrophe bonds over time using the multiple of spread (premium) to expected loss as the key metric. The author noted that this metric was around 5.5 in 2001 – when the average expected losses were under 1%, the average bond spread ws just over 5%, as shown below:

FIGURE 6. Historical Cat Bond Pricing



Source: Lane Financial LLC.

But note a few things about this statistic and how it has varied over time.

- *Multiples Vary by Expected Loss Range* — The multiple really does vary considerably based on the expected loss, as shown earlier in the chart comparing bond spreads compared to the expected loss. For example, based on 2008 issuances bonds with expected losses of around .5% may have had spreads of 5% or so (a multiple of 10) whereas bonds with expected losses of 3% had spreads of 12% or so (a multiple of 4).

- *Speads Vary by Peril and Geography* — The spreads also vary by peril and geography, with Florida hurricane capacity being the most expensive relative to the expected loss.

- *Expected Losses Have Generally Increased* — Catastrophe models have changed over the years and the expected losses for identical exposures have changed as well. Even with multiples declining, overall spreads may not have declined as much as it might appear.

Pricing for Diversifying Peril

One point of comparison discussed in the paper is the pricing of a Mexican earthquake bond to prior earthquake bonds, with the former priced at a much lower multiple of expected loss (2.5 vs. 4.4–7.5). This may be because the peril and geography are attractive to investors, but could also have been a function of the Mexican bond being more recently issued (at the time the paper was written) than the other reference bonds.

Pricing Relative to Corporate Bonds

Yields on catastrophe bonds have been higher than for comparably rated corporate bonds. In the case of BB-rated bonds, the author indicates that catastrophe bond spreads were 2-3 percentage points higher (this differential has varied over the credit cycle and was much higher than this at some points).

Regulatory Considerations

The author briefly mentions prior concerns regarding regulatory issues that were thought to limit the growth of the catastrophe bond market, but notes that these concerns are unfounded.

For example, it is common to use offshore domiciles (Bermuda, Cayman Islands, Dublin) for the special purpose reinsurers rather than onshore (U.S.) domiciles to avoid a variety of regulatory (and tax) implications. The author argues that this has tended to reduce transaction costs and increase efficiency.

In addition, there have been questions raised in the past regarding the regulatory treatment on non-indemnity catastrophe bonds. However, as in the case of ILW's, index-based bonds routinely contain an ultimate net loss provision that has been effective in ensuring the transactions are recognized as reinsurance.

Nonetheless, the author would like to see a more flexible regulatory approach and more clarity on the statutory treatment of risk-linked securities in general.

Tax Issues

The use of offshore domiciles is efficient from a tax perspective because the special purpose reinsurers do not incur any corporate income taxes and are not subject to U.S. taxes so long as they are not deemed to be engaged in a U.S. trade or business (effectively, this means that all negotiations between the insurer and the special purpose reinsurer have to take place offshore). In addition, the insured is able to treat premium payments as a deductible expense in the same way that it would any other reinsurance premiums. As a result, the key tax considerations are those that affect investors in the bonds. In the case of U.S. investors, despite the fact that the investments are structured as bonds, interest income is generally taxed as if it were a dividend.

Securities Laws

Debt securities can be issued in the public or private markets, the choice of which significantly impacts the costs, timing and level of disclosure required — largely due to the SEC's registration process. As a result, catastrophe bonds are sold in the private markets through private placements to accredited or qualified investors only. This limits the amount of publicly available information about catastrophe bonds and reduces the ability of academics to study the market.

Additional Issues

The author lists several factors that could facilitate further growth in the market for risk-linked securities:

- Improved reporting of insured losses could facilitate the creation of more robust loss indices. Currently PCS indices in the U.S. are widely used, but outside of the U.S. comparable loss indices do not exist (some are currently being developed).
- Regulatory capital requirements should acknowledge the counterparty credit risk associated with reinsurance recoveries, which would give a boost to catastrophe bonds which are fully collateralized.
- Personal lines insurance rates should be deregulated and more credit should be given to insurers who lock-in multi-year reinsurance coverage.
- ERISA rules impacting catastrophe bonds should be explored.

Practice Questions

Question 1. Catastrophe bonds can be thought of as primarily a mechanism to allow investors to provide reinsurance to ceding companies. Nonetheless, there are five ways that catastrophe bonds may differ from traditional reinsurance. What are these?

Solution. The five key differences are:

i. Collateral — The policy limit is collateralized and so the insured does not face the credit risk of the reinsurer, which could be very important for very large catastrophes (except in cases such as those described below).

ii. Multi-Year Coverage — Catastrophe bonds are typically issued for multiple risk periods (3 year periods). This locks in pricing and also serves to amortize the significant fixed expenses associated with issuing the bonds (due to banking fees, legal costs, rating agency costs, risk modeling costs, etc. which can amount to more than \$2 million up-front and \$250,000 per annum).

iii. Risk-remote Layers — Catastrophe bonds are typically issued for more risk remote layers (roughly in the range of 1% - 5%).

iv. Indemnity vs. Index Basis — Some catastrophe bonds cover insured losses on an indemnity basis, which reimburses the sponsor for actual losses incurred. However, the majority of catastrophe bonds issued to date have actually provided coverage on an index basis, with payouts determined based on aggregate industry losses, modeled losses or parametric factors.

v. Pricing — Because the investors are pension funds and hedge funds without any natural exposure to insurance risks, it is *arguably* possible for the pricing to reflect a lower risk margin over and above the expected loss than in the traditional reinsurance market (in practice, this is not currently the case). In addition, investors see these as a pure play bet on catastrophe risk, since the bonds are not issued directly by the insurer and are segregated from the rest of its business. They only face the catastrophe risk and not the general business risk of the insurer.

Question 2. Describe the role of the total return swap (TRS) counterparty in catastrophe bond transactions.

Solution. The TRS counterparty provides a fixed-for-floating interest rate swap to the Special Purpose Reinsurer so that investors can receive a LIBOR-based coupon payment despite the fact that the SPR's assets may be invested in fixed rate securities. In addition, they provide a guarantee against mark-to-market losses on the assets to ensure that the assets are always worth at least the notional value of the reinsurance (in the event claims must be paid) or the principal on the bonds (in the event the principal must be returned to the investors.

Question 3. List and describe the various payment triggers that are used in catastrophe bonds.

Solution. The various triggers include:

1. Indemnity Trigger — This works just like in traditional reinsurance, covering the actual losses of the insured, subject to the specified retentions and limits per occurrence or in the aggregate. This does not result in basis risk for the insured, but it does complicate the risk modeling for the investors, requiring them to more carefully underwrite the risks and understand the nuances of the reinsurance business.

2. Index Trigger — An index trigger is intended to pay claims based upon an index that is not tied directly to the insured's actual losses. Three specific forms are used.

   a. *Industry index loss triggers* rely on aggregate industry losses from specified events (e.g. hurricane, earthquake) in specified geographies. Within the U.S. the Property Claims Service (part of ISO) indices are commonly used, but other customized indices can be used as well, including specific weightings of various PCS indices.

   b. *Modeled loss triggers* apply an agreed upon catastrophe risk model, using the physical parameters of the actual events, to the insured's exposure data and pay according to the model's estimated losses rather than actual losses.

   c. *Parametric triggers* pay fixed dollar amounts based upon the physical characteristics of catastrophe events, such as wind speed for hurricanes or magnitude of earthquake.

3. Hybrid Trigger — Hybrid triggers simply blend multiple triggers into a single bond.

Question 4. What are the key considerations in determining the loss triggers to be used for a catastrophe bond?

Solution. The choice of loss triggers is a matter of balancing the trade-off between moral hazard, transparency for the investors (which ultimately impacts investor demand, ease of execution, etc.) and basis risk for the insured.

Indemnity triggers generally eliminate the basis risk for the insured, but require much more extensive disclosure to investors, require much more investor due diligence and complicate claim settlement after an occurrence. The latter issue can be important because of the need to ensure that the SPR is on automatic pilot with a contractual maturity date of the bond.

Index triggers are easier for investors to evaluate — they are not subject to the claims handling practices of the insured and therefore the risk analysis is simplified. Index-based bonds are also not impacted by changes in risk exposure, premium volume, etc., so the overall due diligence process is simplified and execution risk is minimized. However, this comes with increased basis risk, since it is possible that the index losses do not mirror the company's actual claim experience.

Question 5. How are catastrophe bonds rated by the rating agencies and what has been the typical rating for these bonds?

Solution. The rating agencies generally apply standards for the probability of default and/or the expected default losses to assign a rating on catastrophe bonds similar to what they have historically used for corporate bond ratings. They rely on the catastrophe modeling firms to provide them with the relevant statistics to do this. While there have been some issuances of investment grade (BBB- or higher) catastrophe bonds in recent years, the overwhelming majority of the bonds have been below investment grade, rated BB or B.

Question 6. Why is it typical for catastrophe bonds to be issued for multiple risk periods, such as 3 years?

Solution. One reason often given is to lock in pricing and capacity, though it must be remembered that these multi-year catastrophe bonds typically have only one limit over the multi-year period so that if a loss event occurs in the first year then the capacity for the subsequent years' coverage is reduced. A more important factor is the desire to amortize relatively high issuance costs which can exceed 3% of the limit once capital raising fees, legal fees, modeling fees and credit rating fees are included.

Question 7. Describe the composition of the typical sponsors and investors in the catastrophe bond market.

Solution. Historically the sponsors have been split between insurers and reinsurers, with minimal direct corporate or government issuance of catastrophe bonds.

As for the investors, dedicated catastrophe bond funds were responsible for 47% of purchases in 2015. Institutional investors purchased 32% of the issues in 2015, with the rest split among mutual funds, hedge funds and reinsurers. This is a notable change from the earlier days when insurers and reinsurers purchased the majority of the catastrophe bonds issued.

Question 8. Other than catastrophe bonds, what other catastrophe risk transfer products were discussed in the Cummins reading?

Solution. Cummins also discussed sidecars, catastrophe equity puts (contingent capital), catastrophe swaps and ILWs.

Question 9. Briefly describe the historical pricing of catastrophe bonds.

Solution. Pricing in the catastrophe bond market has tended to follow traditional reinsurance market pricing, with the spread to LIBOR representing the annual reinsurance premium. Spreading the catastrophe risk more broadly throughout the capital markets, including to many investors with no existing or natural exposure to catastrophe losses, could arguably result in lower risk margins relative to the traditional reinsurance market, particularly with respect to peak zone risks such as FL hurricane or CA quake. However, pricing in the catastrophe bond market has tended to still reflect a risk margin equal to a healthy multiple of the underlying expected loss. In addition, the catastrophe bond spreads have been considerably higher than comparably rated corporate bonds.

Question 10. Cummins has noted that various regulatory, tax and securities law issues have impeded the growth of the catastrophe bond and risk-linked securities markets. What issues does he specifically mention?

Solution. He notes the following:

i. Regulatory Considerations — Cummins would like a more flexible regulatory approach and more clarity on the statutory treatment of risk-linked securities in general.

ii. Tax Issues — The key tax considerations are those that affect investors in the bonds. In the case of U.S. investors, despite the fact that the investments are structured as bonds, interest income is generally taxed as if it were a dividend.

iii. Securities Laws — Catastrophe bonds are generally sold in the private markets through private placements to accredited or qualified investors only. This limits the amount of publicly available information about catastrophe bonds and reduces the ability of academics to study the market.

Question 11. What factors does Cummins suggest could further facilitate the growth of the catastrophe bond market?

Solution. He notes the following in his conclusion:

i. Improved reporting of insured losses could facilitate the creation of more robust loss indices, particularly outside of the U.S.

ii. If regulatory capital requirements acknowledged the counterparty credit risk associated with reinsurance recoveries then fully collateralized catastrophe bonds would get a boost.

iii. Personal lines insurance rates should be deregulated and more credit should be given to insurers who lock-in multi-year reinsurance coverage.

iv. ERISA rules impacting catastrophe bonds should be explored.

Question 12. Compare catastrophe bonds and catastrophe risk swaps in terms of their impact on risk diversification for the party issuing the bond or entering into the swap. Assume that both the catastrophe bond and the catastrophe risk swap use an indemnity loss trigger with the same attachment points.

Solution. In a catastrophe bond, the issuer is obtaining protection against losses that they are already exposed to. This results in cash flows from the catastrophe bond that are perfectly negatively correlated with its existing risk exposures (in the relevant layer).

With a catastrophe risk swap, assuming the same indemnity loss trigger and the same attachment point, there are also recoveries that are perfectly negatively correlated with the loss. However, there are two additional sources of risk that are assumed. The first is the counterparty credit risk. The second is the catastrophe risk exposure assumed in the other leg of the swap.

If you are trying to evaluate which of these offers the better ”risk diversification” you would argue that the catastrophe bond does because it cedes risk that is perfectly correlated with an existing risk (the ultimate in diversification) and doesn't assume any risk.

Note that this question was asked on the 2013 exam (Question 13) but the model solutions suggest that the catastrophe risk swap offered better diversification. I disagree with the model solution.

Selected Old Exam Questions for Part 4

The following questions relevant for this section appeared on the Old CAS Exam 8 from 2000 to 2010 and on the CAS Exam 9 since 2011.

Hull 24	Coval	Hull 25	Cummins Cat
2007 Q13	2012 Q8		2009 Q8
2007 Q32	2013 Q12		2010 Q29
2008 Q14	2014 Q16		2011 Q12
2008 Q16	2015 Q14		
2009 Q13			
2009 Q15			
2009 Q32			
2010 Q10			
2010 Q9			
2011 Q10			
2011 Q9			
2013 Q10			
2014 Q15			
2015 Q12			
2015 Q13			

For some of these questions I have provided the text of the question and an explanatory solution. These were selected either because they are representative of the questions you are likely to be asked on future exams or because they contain an element that is particularly worthwhile to point out. For the other questions, the CAS solutions should be sufficient to confirm whether your answer is correct.

Important Note: The solutions shown here are intentionally detailed. They contain thorough explanations of the concepts and formulas used to reinforce the main points from the readings and provide an additional teaching opportunity. **Actual exam responses should be much more concise than what is shown here, along the lines of what you will see in the solutions that the CAS releases.**

2007 Exam Question 12

You are given the following information a one-year coupon bond.

- The principal is \$1,000
- The coupon rate is 5%
- The market price is \$875
- The probability of default is 10%

The principal and the coupon are to be paid in one year. In the event of default, neither the principal nor the coupon will be paid. The yield on default free government bonds is 8%. Assume interest rates are compounded continuously.

This question has been slightly modified to reflect the presentation on the current syllabus.

Calculate the yield spread on the one-year coupon bond.

The yield spread is the excess (promised) yield over and above the risk-free rate. To calculate the yield, just set the price equal to the promised cash flows and solve for the yield:

$$\$875 = (\$1{,}000 + 50)e^{-y} \Rightarrow y = 18.23\%$$

This gives a spread of $s = 18.23\% - 8\% = 10.23\%$.

2008 Exam Question 14

You are given the following information for yields on different bonds of varying maturities. Although not specified in the original exam question, assume all bonds are zero coupon bonds paying \$100 at maturity, all yields are continuously compounded and the corporate bonds are all issued by the same company.

	Zero Coupon Yields	
Maturity	Risk Free	Corporate
1	7.00%	7.30%
2	7.00%	7.60%
3	7.00%	7.80%
4	7.00%	8.00%
5	7.00%	8.20%

Assume the recovery rate on defaulted corporate bonds is 30%.

a. Calculate the probability of default between years 2 and 4.

The question is asking for the difference between the unconditional probability of default in year 2 and the unconditional probability of default in year 4.

There are two ways to do this, but in the case of zero coupon bonds the easiest method is to simply solve for the unconditional 2-year default probability, Q_2, which sets the bond price (at its actual yield) equal to the expected (risk-free) present value of the payments.

Recall that if the bond does not default it pays \$100 and if it defaults with probability Q it pays 30%(\$100) = \$30. We can solve for Q from the following:

$$100e^{-.076(2)} = [(1 - Q)(100) + Q(30\%)(100)]e^{-.07(2)}$$

Solve for $Q = 1.7040\%$.

Then do the same thing for the four-year unconditional default probability:

$$100e^{-.08(4)} = [(1 - Q)(100) + Q(30\%)(100)]e^{-.07(4)}$$

Solve for $Q = 5.6015\%$.

The question asked for the probability of default between years 2 and 4, which would be the difference or 3.8975%.

b. Calculate the hazard rate between years 2 and 4.

This is just the conditional probability of default by year 4 given that it does not default by year 2. This is simply:

$$\frac{3.8975\%}{1 - 1.7040\%} = 3.965\%$$

2008 Exam Question 16

You are given the following information about Company XYZ:

- Volatility of assets is 60%
- Risk free rate is 5% per annum, continuously compounded
- The amount of debt to be repaid (interest and principal) in 3 months is $2M

a. Calculate the minimum amount of assets that Company XYZ must have in order to limit its risk neutral probability of default to 5%.

Notice that the original question did not specify that they meant the probability to be the risk-neutral probability, but it is indeed what they expected.

Recall that the Merton model can be used to characterize the value of a firm's equity as a call option on the assets of the firm.

$$E_0 = V_0 N(d_1) - De^{-rT} N(d_2)$$

$$d_1 = \frac{\ln(V_0/D) + (r + \frac{1}{2}\sigma_V^2)T}{\sigma_V \sqrt{T}}$$

$$d_2 = d_1 - \sigma_V \sqrt{T}$$

Within this context, the call option will be exercised (the equity holders will repay the debt) if the assets are worth more than the liability. The *risk-neutral* probability that an option will be exercised is $N(d_2)$ and therefore the probability of NOT exercising the option, which is to say that the probability of the equity holders not repaying the debt to buy back the assets of the firm, is $1 - N(d_2)$. This is the risk-neutral default probability.

Using the values we simply set $1 - N(d_2) = 5\%$ and solve for the value of the assets V_0.

Assets	**3,384,159**
Debt	2,000,000
Risk-free Rate	5%
Maturity	0.25
Asset Volatility	60%
d_1	1.9449
d_2	1.6449
$N(d_1)$	0.9741
$N(d_2)$	0.9500
$1 - N(d_2)$	5%

b. Determine the value of the company's equity today using the amount of assets in part (a) above.

The equity is just the value of the call option, so using the same calculations we have $E =$ \$1,420,124.

c. Calculate the expected loss on the debt as a percentage of its no default value.

Since the assets are worth \$3.38M and the equity is only worth \$1.42M, the difference is the market value of the outstanding debt, or \$1.96M. The risk free value of the debt is \$1.975M and so the difference is the expected loss from default, which comes to .56% of the risk free value of the debt.

The full calculations, without rounding, are shown below:

Assets	3,384,159
Equity	1,420,124
Risk-Free Value of Debt	1,975,156
Market Value of Debt	1,964,035
Expected Loss	11,121
Expected Loss as %	0.56%

2010 Exam Question 9

Given the following information:

- Value of a company's assets today is $19,000,000$
- Value of a company's equity today is $5,000,000$
- Risk-free rate is 3% per annum
- Debt of 15,000,000 to be repaid in one year
- Instantaneous volatility of equity =.5
- Volatility of assets = .25

a. Explain what assumptions Merton makes that allow Merton's model to be used in estimating default probabilities for companies.

He treats equity as a call option on the firm's assets with a strike price equal to the face value of the debt and maturity equal to the maturity date of the debt. To make this consistent with the standard Black-Scholes formula, the debt is assumed to be zero coupon debt with a single maturity date.

b. Calculate the probability of default using Merton's method of using equity prices to estimate default probabilities.

Warning: The information in this question was inconsistent and so you get a different answer depending on what calculations you do. The CAS eventually awarded full credit to both solutions.

From the Merton model, we have the value of the firm's equity as:

$$E_0 = V_0 N(d_1) - De^{-rT} N(d_2)$$

$$d_1 = \frac{\ln(V_0/D) + (r + \frac{1}{2}\sigma_V^2)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma_v\sqrt{T}$$

Within this context, the call option will be exercised (the equity holders will repay the debt) if the assets are worth more than the liability. The *risk-neutral* probability that an option will be exercised is $N(d_2)$ and therefore the probability of NOT exercising the option, which is to say that the probability of the equity holders not repaying the debt to buy back the assets of the firm, is $1 - N(d_2)$. This is the risk-neutral default probability.

This means all we need to do is calculate $N(d_2)$:

$$\begin{aligned} d_2 &= \frac{\ln(V_0/D) + (r - \frac{1}{2}\sigma_V^2)T}{\sigma\sqrt{T}} \\ &= \frac{\ln(19/15) + (.03 - \frac{1}{2}.25^2)}{.25} \\ &= .9405 \end{aligned}$$

From this, the probability of default is $1 - N(d_2) = 17.35\%$.

Notice that there was an inconsistency in the question. What would happen if we actually used the information given to not only calculate $N(d_2)$ but also to calculate the entire value of the equity? In this case, we would have the following results from the Merton Model:

N(d1)	0.883
N(d2)	0.827
V	19.00
D	15.00
r	3%
T	1
Equity	4.75

But this isn't the value of the equity given in the question, which was 5 million. This confirms an inconsistency.

Why is this relevant? Because there was another *correct* way to answer this question. Using that method would give a different answer due to the fact that it uses the 5 million value for equity that was given in the question.

Recall that Merton's model is actually used in reality in cases where we do not know either the value of the assets nor their volatility. So we use a relationship between the equity and asset volatilities and values as well as the Black-Scholes equation:

$$E_0 = V_0 N(d_1) - De^{-rT} N(d_2)$$

$$\sigma_E E_0 = N(d_1)\sigma_V V_0$$

Since we were given $\sigma_E, E_0, \sigma_V, V_0$, we could have plugged those into the above formula and solved for $N(d_1) = .5263$. Then we could have used the values given to solve the Black-Scholes equation for the value of $N(d_2)$ as follows:

$$N(d_2) = \frac{V_0 N(d_1) - E_0}{De^{-rT}} = .343$$

Which would have given us a default probability of $1 - N(d_2) = .656$.

c. Calculate the expected loss on the debt as a percentage of its no default value.

If the debt doesn't default its value is $15e^{-.03} = 14,556,683$. However, we know that its current market value is only equal to the difference between the asset value and the equity value, or $14,000,000$ based on the values given in the question (or $14,252,956$ using the alternative value for the equity discussed above).

This gives expected losses in the event of default as the difference in market values, or $556,683$ on a present value basis or $573,637$ on a nominal basis (or lower if the alternate calculations are used). As a percent of the no-default value, this is 3.82% on a discounted basis and 3.94% on a nominal basis.

2011 Exam Question 9

Given the following information for Bonds A and B, **both of which are zero coupon bonds**:

	Bond A	Bond B
Par Value	1,000	1,000
Term (Years)	2	3
Yield (continuous)	6.00%	???
Recovery Rate	20%	20%

And the following additional assumptions:

- Default is possible only at the end of the year in which the par value is due (i.e. just prior to maturity).
- The unconditional default rate is constant for years one and two and is the same for both bonds.
- The unconditional default rate for year three is 2%.
- The yield on similar risk-free bonds is 5%, continuously compounded, for both maturities.

Calculate the implied continuously compounded yield for Bond B.

Note that the original exam question did not specify that these are zero coupon bonds. You could have made any assumption you wanted to make about the coupon rate, but obviously it is easiest assuming no coupons at all. In addition, since I am going to assume these are zero coupon bonds, I am also going to assume that there's no "default" until maturity. The original question only said that default occurs at the end of the year.

This is a simple version of the default rate calculations in Hull except that you need to do it once to get the value for the year one and year two annual default rates, Q, using the information for Bond A and then do it in reverse to get the yield on Bond B.

Begin with Bond A. If this were a risk free bond, it's value would be:

$$G = 1,000e^{-.05(2)} = 904.837$$

Given that it has default risk, the yield is higher and the price is lower:

$$B = 1,000e^{-.06(2)} = 886.920$$

The difference in these two prices is the expected loss from default, or $G - B - 17.917$.

Another way to calculate the expected loss from default is to recognize that the bond can only default at maturity and the probability of this occurring is $Q(2) = Q_1 + Q_2 = 2Q$ since the year one and year two annual default rates are assumed to be equal. The loss given default is 80% of the face value, or \$800, so the expected loss from default on a present value basis is:

$$\$800(2Q)e^{-.05(2)}$$

Setting this equal to 17.917 we can solve for $Q = 1.238\%$.

Now we can use the probability of default through year 3, $Q(3) = 1.238\% + 1.238\% + 2\% = 4.475\%$ to get the expected loss from default of Bond B:

$$\$800(4.475\%)e^{-.05(3)} = 30.815$$

If Bond B were risk free its price would be $G = 1,000e^{-.05(3)} = 860.708$ and so we know that the price of the bond is $B = 860.708 - 30.815 = 829.893$. From this we can get the yield as:

$$829.893 = 1,000e^{-y(3)} \Rightarrow y = 6.22\%$$

I mentioned above that the actual exam question didn't specify the coupon rate. If you did assume a coupon rate, then you would have to be more careful about how much is lost at the time of default. In the textbook, the example given with coupon bonds determined the loss at the time of default as the difference between the risk-free value of the remaining cash flows *at that time* and the recovery amount (\$200 in this case).

Notice as well that I added text to the question indicating that the bonds could only default just prior to maturity. This is because in practice, a bond issuer cannot or would not default before they actually owed anything. So if the bonds don't pay coupons, there would be no reason for them to default until maturity. If instead, as the CAS sample answers did, you wanted to assume that default could occur at the end of any year, then the calculations would be a bit more involved. You'd have to calculate the expected loss from default assuming that at year 1 the loss given default is the risk-free value of a bond with one more year to maturity ($1,000e^{-.05(1)} = 951.229$) less the \$200 recovery, or \$751.229. To get this amount, I had to assume that the risk-free rate was 5% for both the one-year and two-year maturity since the discount rate used to get the value in one year is the current forward rate.

Then discounting this to the present and multiplying by Q would give you one portion of the expected loss from default. The other would be the expected loss if default occurred at year 2, which is \$800 again before discounting and multiplying by Q. Done this way, $Q = 1.245\%$. Similar calculations would be done for Bond B, allowing for default to occur at year 1, year 2 or year 3.

2015 Exam Question 14

A country recently began allowing financial institutions to package catastrophe bonds into collateralized debt obligations (CDO) that allowed tranches to be constructed based on the performance of the underlying catastrophe bonds.

Critics objected that allowed catastrophe bond CDO tranches with AAA ratings could lead to a repeat of the subprime mortgage crisis.

a. Discuss three ways the risks associated with catastrophe bond CDOs differ from the collateralized mortgage obligations constructed before the subprime financial crisis.

Some differences include the following:

(Note that I will provide more than the three asked for in the question)

- When subprime mortgages were pooled, the intent was to create pools of largely uncorrelated underlying mortgages, such as by ensuring a minimal amount of geographic overlap. However, it was not possible to know for sure how correlated subprime mortgages in Las Vegas were with subprime mortgages in Florida. In contrast, if catastrophe bonds are used as the underlying collateral in a CDO, it would be easier to know how uncorrelated the underlying risks were, since earthquakes and hurricanes in sufficiently diverse geographic regions are reliably uncorrelated.

- Because of the strong incentives for banks, rating agencies and investors to "create" subprime mortgage-backed CDOs, a substantial portion of subprime mortgages issued just prior to the financial crisis were either highly likely to default (e.g. loans to borrowers with no income, no job and no assets) or were highly dependent on a robust market for new loans (e.g. highly dependent on refinancing of unaffordable mortgages upon rate resets). In contrast, it is far more difficult to create catastrophe bonds that are highly likely to be triggered since they require the occurrence of a natural catastrophe such as a hurricane or earthquake. This lack of moral hazard is a material difference between catastrophe bond CDOs and CDOs comprised of subprime mortgages.

- Rating agencies, in their effort to be accommodating to banks and earn fees, were able to manipulate the correlation assumptions and the home price appreciation assumptions that were so critical in establishing the ratings for the AAA tranches of subprime mortgage CDOs. In contrast, catastrophe bond modeling assumptions are arguably more grounded in scientific modeling of the underlying loss events, and thus arguably less susceptible to manipulation.

- The risks in subprime mortgages were closely tied to changes in interest rates and other market variables that made their underlying risks more systematic and less diversifiable, whereas the underlying catastrophe risk is more purely diversifiable for capital markets investors.

Discus two potential challenges arising from widespread construction and investment in the catastrophe bond CDOs that could contribute to a future financial crisis.

One concern would be whether synthetic CDOs would be created that allowed bonds with exposure to the same underlying loss event to be put into different CDOs, thereby magnifying the aggregate amount at risk beyond the value currently at risk. For instance, a single building in Miami can be insured only once and so exposure to a loss from a catastrophe event can only exist in the reinsurance market or in a catastrophe bond once. But if, as was the case with subprime mortgages, investors were able to take synthetic exposure to the loss on that one building, it would be possible for many multiples of the total value of the building to be at risk across multiple bonds. This occurred with subprime mortgage CDOs, which magnified the impact from the collapse of the subprime mortgage market.

Another concern would be that highly rated tranches of catastrophe CDOs, if they existed, might be used as collateral for other risky investments. Given the potential for catastrophe events to occur without notice and for the payouts on the bonds to occur quickly, such use of CDO tranches as collateral for other risky assets could result in liquidity problems that could bleed into other markets.

A final concern (a third, even though the question asked for only two) is that widespread creation of catastrophe bonds could result in CDOs being formed with more correlated exposures, making the senior tranches more likely to suffer losses.

Propose and briefly explain two possible regulations for construction or rating of the new catastrophe bond CDOs that could help alleviate some of the challenges in Part b above, without greatly discouraging innovation and investment.

One regulation that would alleviate some of the challenges is to require a minimum level of diversification across geographic regions and perils. Another would be to more carefully define the criteria for the CDO ratings, including modeling standards and definitions of what is required to achieve a given rating. And another would be to recognize the unique liquidity risk in catastrophe bonds and not allow them, or CDOs based on them, to be used as collateral for other risky investments.

Note that the sample solutions provided by the CAS for this question contain a long list of excellent observations about the challenges (Part B) and proposed regulations (Part C) related to catastrophe bond CDOs. However, Part B specifically asked about the challenges that could lead to a "financial crisis". It doesn't seem to me that many of the points made, while valid in general, get to the heart of what could trigger a crisis. Similarly, the sample solutions make multiple references to the collateral in a catastrophe bond. However, this impacts the insured mostly and not really the investors, except to the extent (as happened in the Lehman CDOs) that the collateral could be lost as the result of a risky investment and not just because of the occurrence of a catastrophe event.

Appendix: Normal CDF Table

TABLE 1. Cumulative Normal Distribution (Positive x)

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 2. Cumulative Normal Distribution (Negative x)

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Part 5

Financial Risk Management (b)

Part 5 Table of Contents

Stulz: Rethinking Risk Management

Why Firms Manage Risk

Firms can maximize their value through the use of derivatives to hedge certain financial risks, thus minimizing the volatility of their cash flows and minimizing the costs of financial distress.

Market Efficiency and Risk-Taking

Recall from the discussion of market efficiency in BKM Chapter 12 that modern finance theory suggests that attempts to earn higher returns in financial markets is likely to require bearing additional risk. As a result, it is inappropriate for firms to attempt to profitably trade in interest rates, currencies or commodities unless they have a comparative advantage bearing these risks (see below). In addition, firms that already have a natural exposure to these risks may find it desirable to hedge these risks.

However, keep in mind that modern finance theory suggests that much of this risk will be naturally diversifiable by the shareholders. As a result, the firms themselves may not necessarily have to hedge these risks, unless strategies designed to minimize cash flow variability result in the elimination of real costs.

Financial Distress Costs

A firm with highly variable cash flows may face real costs as a result of their cash flow variability and therefore may be able to reduce these costs through risk management. The three examples of costs that can be reduced through risk management are:

- Reduce Bankruptcy Costs — If wild swings in cash flow make it difficult to meet debt payment obligations, the company may be forced to file for bankruptcy. This will involve direct costs (lawyers) as well as indirect costs associated with the interference with the normal business operations and investments. But even short of bankruptcy, firms facing high risks of bankruptcy may find it difficult to raise funds for new investments, leading to an *under investment problem* whereby profitable projects have to be passed up.

- Reduce Payments to Stakeholders — As noted above, shareholders are likely able to diversify their risks more efficiently than the firm can. But other stakeholders, including managers, employees, suppliers, etc. have much more non-diversifiable firm specific risk and they will require larger payments as risk increases. For example, an employee of firm with a high probability of failure will likely demand higher wages than if the firm were stable.

- Reduce Taxes — Increasing marginal tax rates, limits on the use of tax loss carry forwards and the alternative minimum tax all serve to make extreme levels of profit (too high or too low) less desirable than a more stable income level.

Primary Objective of Risk Management

The author argues that the primary objective of risk management is to eliminate costly lower tail outcomes, thereby minimizing the likelihood of financial distress and preserving the financial flexibility to carry out their investment objectives.

Risk Management in Practice

The theory presented above would seem to suggest that smaller firms would hedge more often than larger firms and firms that hedge would do so primarily to reduce volatility. In practice though, surveys find that the opposite is true — large firms tend to hedge more than small firms (perhaps due to the investment in personnel, training and computer systems required) and often firms engage in selective hedging, whereby they allow their views of the markets to influence whether or not, and to what degree, they hedge (according to results of surveys conducted by Wharton-Chase).

Other survey results of note include:

- Most hedging seems to be for executed transactions and near term exposures, rather than company-wide financial risks.

- Firms that hedge report hedging more of their risk if they think the market is more likely to move against them than move in their favor — that is, they want to avoid large losses but not at the expense of giving up large gains.

Two Case Studies

The experiences of Metallgesellschaft and Daimler-Benz are summarized. The key elements of these cases can be summarized as follows:

Metallgesellschaft

In the case of Metallgesellschaft, their risk management arm appears to have over hedged their position in order to profit from their expectation that oil futures would continue to exhibit backwardation. The term backwardation used here refers to the situation where the futures price is below the expected future spot price. As time passes, a long position in the futures contract is expected to earn a profit as the futures price converges to the expected spot price (recall as the expiration date approaches the futures price and the spot price have to converge). Since they expected to roll over short term hedges, they expected to be able to earn these profits from the rise in the futures price.

When the spot oil prices fell, their initial position (the long term oil sale agreements with their customers) experienced large paper profits and their hedge position experienced huge paper losses, as expected. But because the losses had to be marked to market while the gains did not, the hedge looked like a speculative loss. With the overhedge, the losses were even more severe and the firm was forced to terminate their hedge.

Daimler-Benz

Daimler-Benz lost DM1.56 billion on a mostly dollar denominated order book due to changes in the dollar-DM exchange rate that were not hedged because management's view was that the exchange rate would remain above 1.55 DM/$.

Comparative Advantage

One of the critical points made in this paper is that firms may in fact acquire valuable knowledge about certain financial markets in the course of their business and it would be acceptable for them to exploit this knowledge by adjusting their degree of risk management. The author refers to this as *speculative hedging*.

The key danger in speculative hedging is that the firm's knowledge may not in fact offer a comparative advantage.

Worse, the firm may not even fully understand where its comparative advantage comes from and may engage in activities that fail to capitalize on the true source of their advantage. The example provided is that of a foreign exchange (FX) trading operation. It is argued that a large FX operation may in fact have a comparative advantage in trading currencies, but the advantage does not come from the superior views of their trading staff. Instead, the advantage comes merely from their size — allowing them to take offsetting positions and profit from market-making (i.e. they earn their superior returns by earning the bid-ask spreads on positions that net out to zero).

Capital Structure

The capital structure of a firm reflects the mix of debt and equity to finance the firm's operations. As the ratio of debt to equity rises, the risks of and costs of financial distress grow. But because we have argued that the goal of risk management is to minimize financial distress costs, it should be clear that risk management can be a substitute for more equity capital in the capital structure. The more a firm hedges its financial risks, the less equity it needs (or conversely the more debt it can carry).

Three situations are discussed:

- *Highly rated firm with low debt to equity ratio*

  A firm like this has very low probability of financial distress and therefore can afford not to hedge the financial risks it assumes in the normal course of business. It may even be able to take speculative positions in financial markets if it had a comparative advantage.

But the key issue here is that a firm like this might even want to consider increasing its debt ratio and substituting risk management for equity capital. This allows it to maintain the same probability of financial distress, but also benefit from the advantages of debt (lower taxes, strengthen management incentives, increase concentration of equity ownership).

- *Firm with low credit rating and significant probability of financial distress*

  This firm should engage in active risk management to reduce the probability of financial distress.

- *Firm in distress*

  This firm is already in distress and therefore should not hedge its risks because most of the gains will go to the debtholders anyway. In fact, this firm should aggressively take on more risk, since the downside is borne by the debtholders and the upside might be large enough to get the firm out of distress — a windfall for the shareholders.

Management Incentives

A common problem for publicly traded firms is that often the managers of the firms act in their own interests and not necessarily in the interest of the shareholders. A study by Peter Tufano highlights this problem with respect to decisions of firms to hedge financial risks. He notes that the only important determinant of the extent to which various gold firms hedged their gold price risk was the extent to which the managers were also shareholders. The greater management's share ownership, the more they hedged. In addition, compensation in the form of stock options tended to result in less hedging, since the option holder benefits from volatility.

Recognizing that management may have their own interests in mind when they determine the degree to which they hedge risks is important. For example, the corporate treasury department may have a strong incentive to take risky bets in an effort to stand out if things go well.

Value at Risk

In modern financial theory, risk is often measured using variance of returns. Using this definition of risk, risk management might be thought of as minimizing the variance of returns. However, this paper shows that many firms don't actually define risk in this fashion and instead tend to think of the goal of risk management as the avoidance of *lower tail outcomes*, thus making variance a somewhat inappropriate risk measure.

An alternative risk measure, known as Value at Risk or VaR, attempts to quantify the probability of these lower tail outcomes and has become quite commonly used in banking and financial firms for short-term risk measurement. Students should have been exposed to VaR on the preliminary exams, but a short review will be useful because the VaR concept is referred to in most of the subsequent readings in this section.

VaR Background

The Value at Risk for a given portfolio of financial assets is a measure that attempts to make a statement of the following form:

> "We are X percent certain that we will not lose more than V dollars in the next N days."

VaR is simply trying to answer the question, "given the current portfolio, how bad can things get in N days? How much can the value decline within a stated time interval?"

When measuring VaR, it is not practical to try to model every conceivable event that could affect the value of a given portfolio. Therefore, statements about VaR are generally limited to the impact of key economic variables such as interest rates, exchange rates, volatility, etc.

Consider a simple example. The portfolio contains \$10 million of a single stock and the goal is to find the 99% confidence level for losses over 10 days. In the VaR definition from above, $X = 99$ and $N = 10$. They further assume a daily volatility of 2% of the asset value, or 2%(\$10 million) = \$200,000. By further assuming that each day is independent, then the 10-day volatility is equal to:

$$\sigma_{day}\sqrt{10} = \$632{,}456$$

Then, we typically make two further assumptions to simplify the formulas. The first is that the expected daily price changes are small enough (especially relative to the standard deviation) that they can be assumed to be zero over a small period of time. The second assumption is that the change is normally distributed. So from these assumptions, the 10-day price change is normally distributed with a mean of approximately zero and a standard deviation of \$632,456.

A normally distributed random variable will have a value that is less than 2.33 standard deviations below its mean only 1% of the time. We know this because $N(-2.33) = .01$. As a result, the probability is only 1% that the value of this portfolio will decrease by more than 2.33(\$632,456) = \$1,473,621 during this 10-day period. Therefore, the 99% 10-day VaR is \$1,473,621.

Limitations of VaR for Risk Management Purposes

The author argues that VaR is not that useful over the longer time horizons of interest to most firms, for two reasons:

- In theory, a one-year VaR could be calculated. But since the goal is to find the extreme events that occur say 1 time out of 100, you would not be able to test whether the results are reasonable or if the model is properly calibrated because you couldn't possibly collect data over enough unique one year periods.

- VaR typically assumes normal distributions whereas real world data, especially over long time horizons, has much fatter tails than the normal distribution would suggest. This isn't a flaw in the VaR concept, but it does mean that many commonly used formulas for computing VaR may be inadequate.

Cash Flow Simulations

As an alternative to VaR, the author suggests performing cash flow simulations (e.g. Dynamic Financial Analysis), which allow you to measure risk over longer time horizons and study the path of firm value over the entire period and not just at the end of the period. It also has the flexibility to allow for non-normal distributions and correlations among variables.

Managing Risk Taking

Given that it might make sense for firms to take risks to the extent that the expected gains from the activity exceed the costs associated with the increased risk of financial distress, it is important to create the proper incentives for management.

The author has two specific recommendations:

- Gains from risk taking activities must be measured on a risk-adjusted basis.

- Compensation systems should be designed so that managers are not compensated merely for taking risks. They should be compensated only for earning excess risk adjusted returns. That is, they should only be compensated for increasing shareholder wealth.

Practice Questions

Question 1. The text states that reducing cash flow variability caused by exposure to financial risks may not be an appropriate goal for most firms. Why?

Solution. The primary reason given is that many of the risks that result from financial exposures are likely to be more efficiently managed by the firm's shareholders than by the firm itself. The firm should engage in active risk management of these risks only if it can reduce real costs.

Question 2. What three real costs might a firm be able to reduce through financial risk management?

Solution. The three real costs discussed are:

- Bankruptcy Costs — direct costs including legal expenses and indirect costs including loss of flexibility or the inability to raise funds for new investment;

- Payments to Stakeholders — including employees (who might demand higher wages for the risk of losing their jobs) or customers who might be reluctant to do business with the firm if there's a risk that the products will not be supported;

- Taxes — Extreme volatility of earnings can result in higher taxes due to the convex nature of the tax rates.

Question 3. Surveys conducted regarding the use of risk management have uncovered several interesting observations about real world hedging activities of firms. List some of these.

Solution. Some of the observations were:

- Large firms hedge more than small firms, even though smaller firms most likely have the more volatile cash flows;

- Firms often engage in selective hedging based in part on their own views of risk in different markets;

- Most firms hedge the risks associated with executed transactions or near-term exposures, not companywide risks;

- Firms tend to shy away from having to give up gains in order to avoid losses, unless they are concerned about suffering a very large loss.

Question 4. Stulz suggests that it might be appropriate for real world hedging to be more selective and consistent with the survey observations. What reason does he give?

Solution. Firms may gain comparative advantages through the course of their normal business activities that can and should be used to determine whether and how much risk to hedge. Of course, he warns that it will always be difficult to know whether the comparative advantage is real or even if the source of the advantage is understood well enough by the firm to properly take advantage of this knowledge.

Question 5. Stulz argues that highly rated firms might be able to afford not to hedge their financial risk assumed in the normal course of business because their strong capital base gives them ample protection against the risks and costs of financial distress. What reason does he give for why these firms might hedge their risks anyway?

Solution. By hedging their financial risk, they may actually be able to reduce the amount of equity capital in the firm and gain more of the advantages inherent in the use of financial leverage (among other benefits, the use of debt reduces taxes and strengthens management incentives).

Question 6. Stulz gives two reasons for why management may favor a different degree of risk management than the firm's shareholders. Give a reason why they may want to hedge more and why they might want to hedge less.

Solution. The two reasons are:

- Hedge More — It is argued that the firm's shareholders are likely to own a diversified portfolio of shares and therefore the risks of any one firm are naturally managed through diversification. However, the managers of the firm are unlikely to be as diversified — they likely have a relatively large proportion of their wealth tied up in the firm's stock, options on the firm's stock and future compensation. So the managers will be more concerned with the specific risks of the firm than the typical shareholders are.

- Hedge Less — To get ahead in a firm, it is often desirable to stand out. One way to do so would be to take risky bets and hope that they turn out favorably. If they do, the management is likely to be rewarded; if they don't, it's likely to be the shareholders who feel the penalty. This asymmetric benefit to risk taking creates a strong incentive to take risk.

Question 7. What does Stulz think should be the goal of risk management? Why?

Solution. Stulz believes that the goal should be to *eliminate lower tail outcomes* so that firms can lower the expected costs of financial distress but at the same time preserve the firm's ability to exploit any comparative advantage in risk taking it may have.

Question 8. Peter Tufano studied the hedging behavior of 48 publicly traded gold mining companies. What did he find as the only systematic determinant of the firms' hedging decisions? What rationale does Stulz give for this result?

Solution. The degree to which managers of the firms owned the shares of the firm — the larger the managers' equity stakes, the more they hedged. The rationale is that these managers are not as diversified as other shareholders (their major source of wealth and future income is directly correlated to the share price of the firm) and hence they behave in their own interests by hedging more.

It is interesting to note that when managers' stakes consisted more of options than actual shares, the firms tended to hedge much less or not at all. This is because managers' options have one-sided payoffs and thus benefit from increased volatility.

Question 9. When firms hedge more simply because the managers' own interests differ from those of the shareholders, this is likely to hurt the shareholders. Why does Stulz argue that the increased hedging by firms with more concentrated managerial ownership may actually be value enhancing for the shareholders?

Solution. Stulz argues that firms with large financial exposures that are outside of the control of managers may find that the managers have a disincentive to own shares. But allowing the managers to hedge their risks more can offset this disincentive, causing them to own more shares in the firm and thus allowing the shareholders the additional benefit of stronger managerial incentives to focus on increasing the value of the firm.

Question 10. Stulz argues that Value at Risk (VaR) risk measures cannot be used to facilitate risk management, which is focused on the elimination of lower tail outcomes over an extended time period, in part because VaR is intended to be used over much shorter time horizons. What are the two major difficulties with extending VaR over longer time horizons?

Solution. Since VaR focuses on outcomes in the 99th percentile, longer time periods mean that there is substantially less data available to estimate model parameters or test the model. Secondly, VaR as it is traditionally implemented relies on normal distributions.

Question 11. What alternative to VaR does Stulz prefer and what does he say are some of the advantages?

Solution. He prefers Cash Flow Simulations that can identify if and when firms fall below their target and into the range of financial distress (e.g. when their credit rating falls below some level). This allows them to see the path of firm value during a period of time, rather than just the distribution of ending values. It also facilitates sensitivity analysis and Monte Carlo simulations that can easily incorporate non-normal distributions and serial correlations.

Question 12. Stulz argues that once a company concludes that a transaction's expected profit exceeds the increase on its expected financial distress costs they should then assess actual performance on a risk-adjusted basis. What method does he suggest can be used to do this?

Solution. When evaluating actual returns, they should look only at the abnormal returns over and above the returns on investments with comparable risks. Further, to the extent that the transaction increases the probability of financial distress, they should measure the normal return relative to a larger capital base that includes the additional capital needed to retain the same probability of financial distress.

Culp, Miller & Neves: Value at Risk — Uses and Abuses

Development of Value at Risk (VaR)

VaR first emerged as a means for trading firms to measure and aggregate risks in their different trading portfolios. It had several features that were of particular value to trading firms:

- Consistent — Expressing the risk in dollar terms made it easy to use across different lines of business and products;
- Probability Based — Management can specify the confidence level and use VaR as a forward looking measure of market risk;
- Common Time Horizon

Measuring VaR

A basic example of the VaR concept for a stock portfolio was provided in the Stulz reading notes, so I won't dwell too much on the discussion in Culp of the calculation details (most of which is in the Appendix and therefore not on the syllabus).

The main points about the calculation of VaR that are made by the authors include:

- VaR is a two-step process. First, generate the probability distribution for the price or returns of each security. Second, aggregate the individual distributions using an appropriate assumption about their correlations.
- Traditional VaR measures usually assume that the portfolio composition does not change over the time horizon. This is fine for very short time horizons, like 1-day or 5-day, but for longer time horizons this can be problematic.

Uses of VaR

Since its inception, VaR has been used extensively by trading firms. Some of its uses included:

- Risk Reporting — Its simplicity allowed for quick and efficient risk reporting to senior management.
- Risk Control — It also proved useful as a means to monitor and set risk levels by market, by trading group or by counterparty.
- Risk Management — VaR provided useful information for hedging risks and evaluating any particular transaction's effect on portfolio risk.

- Capital Allocation — VaR has also been useful to measure risk adjusted returns for compensation purposes or for capital allocation purposes.

- Exposure Monitoring — Non-trading firms tend to use measures like VaR more for monitoring external fund managers or assessing collateral requirements for customers.

Risk Management Objectives

Firms can either be *value risk managers*, concerned with total firm value at any point in time, or *cash flow risk managers*, concerned with cash flow variability.

Four Derivatives Disasters

In the early 1990's, several high profile derivatives disasters occurred. In each case, the entities involved suffered tremendous losses in their portfolios (mostly from derivatives) which some have argued could have been avoided had the firms been using risk measures such as VaR. As the authors discuss though, simply measuring VaR is not a sufficient element of risk management.

Proctor & Gamble

In 1993, P&G entered into a fixed for floating interest rate swap to reduce its financing costs significantly but in exchange they also wrote an interest rate put to its swap counterparty.

Although you do not need to know this for the exam, I think it would be helpful for you to know the details of the swap. The term of the swap was 5 years, the notional amount was \$200 million, P&G agreed to receive a fixed rate of 5.3% from Bankers Trust and pay a floating rate equal to the (average) 30-day commercial paper rate less 75 basis points, plus a *spread*. Ignoring the spread, this allowed P&G to potentially save a lot on their financing costs — essentially 75 basis points per year if we assume that the 5.3% fixed for 30-day CP rate was itself a fair exchange.

The complication was that the spread was determined as follows. At inception the spread was equal to zero. Then, on the first floating rate reset date in May 1994, the spread would be calculated according to the following formula:

$$\text{Spread} = \max\left[\frac{\frac{\text{5-Yr UST Yield}}{5.78\%}(98.5) - \text{30-Yr 6.25\% UST Bond Price}}{100}, 0\right]$$

At first glance, this looks like it might be related to the difference between the 5-year and the 30-year yields. And at the yields as of the inception date in late 1993, the spread payment was zero and would be zero for very small changes in yields.

However, if you look more carefully at this, it is really the difference between the *yield* (of the 5-year bond) and the *price* of the 30-year bond. Since yields and prices move in the opposite direction, this is really a bet that rates will not rise. If the 5-year yield rises, the first term increases by a multiple of the change in the yield, where the multiple is $.985/.0578 = 17.04$. At the same time, if the 30-year yield rises, its price falls.

We can use the modified duration of the 30-year bond to determine how much the price would change for a rise in the yield, which turns out to be something on the order of 13.68 times the increase in the yield. Therefore, if both yields rise, the spread rises by approximately:

$$17.04(\text{Change in 5-Year Yield})\ \textit{plus}\ 13.68(\text{Change in 30-Year Yield})$$

This is very sensitive to changes in the yield.

For example, if interest rates rose by just 1% for both the 5- and 30-year securities (from their then current levels of 5.02% and 6.1% respectively), each of P&G's semi-annual spread payments for the remaining 9 payment dates would total approximately $13 million!

What actually happened? It turns out that in January 1994 the deal was renegotiated slightly and then by March 25, 1994, still two months before the reset date, rates had already risen by more than 1%. At that point, the undiscounted value of the 9 spread payments was over $100 million. Given the magnitude of the loss, the contract was liquidated prior to the actual reset date with a loss to P&G of over $100 million. It is interesting to note that had the deal continued until the May reset date, the losses would have been more than twice this amount.

Table 1 on the following page shows the spread payments (the simple sum of all 9 payments) using actual 5- and 30-year yields and prices (note that I do not net out the 75 basis points savings).

This numerical example shows how dangerous leveraged derivative instruments can be. In order to save 75 basis points per year on a $200 million debt instrument, P&G took a bet on interest rates that could have cost them close to $300 million! This is why Warren Buffet referred to *certain* derivatives as "financial weapons of mass destruction".

So would a VaR system have prevented this? *Possibly*.

If the risk that was taken by writing the put was better understood by senior management, and if senior management was really monitoring individual transactions' VaR, then perhaps the deal would not have been approved. But it should be noted that VaR was never intended to be a transaction specific measure — most of its appeal came from its ability to aggregate across different markets and risks. To use VaR for individual transactions, P&G would have had to measure the total VaR of their swap portfolio and measure the additional VaR created by this deal. And while this might be a good idea for some firms, note that VaR is a value risk measure while most of P&G's treasury risk management activities most likely involved cash flow risk management and so it is not clear what a measure like VaR would have told them.

Barings

Nick Leeson, a young trader at Barings Bank made enormous bets on the Japanese stock market using futures contracts and lost over $1 billion. Clearly, a VaR measure would have told the senior management to shut down this operation much sooner, except that it is unlikely that

TABLE 1. P&G Leveraged Swap Spread Payments

Date	5-Yr CMT Yield (%)	30 Year Bond Clean Price	Spread (%)	Undiscounted Spread ($ millions)
11/5/1993	5.03	100.6	0.0%	0.0
11/12/1993	5.04	101.4	0.0%	0.0
11/19/1993	5.04	98.8	0.0%	0.0
11/26/1993	5.13	100.1	0.0%	0.0
12/3/1993	5.14	100.1	0.0%	0.0
12/10/1993	5.10	100.7	0.0%	0.0
12/17/1993	5.18	99.6	0.0%	0.0
12/24/1993	5.16	100.5	0.0%	0.0
12/31/1993	5.14	98.8	0.0%	0.0
1/7/1994	5.21	100.3	0.0%	0.0
1/14/1994	5.03	99.4	0.0%	0.0
1/21/1994	5.06	99.6	0.0%	0.0
1/28/1994	5.05	100.4	0.0%	0.0
2/4/1994	5.14	98.7	0.0%	0.0
2/11/1994	5.36	98.0	0.0%	0.0
2/18/1994	5.40	95.2	0.0%	0.0
2/25/1994	5.60	94.1	1.3%	11.7
3/4/1994	5.74	92.6	5.2%	47.1
3/11/1994	5.85	91.8	7.9%	70.9
3/18/1994	5.91	91.7	9.0%	81.0
3/25/1994	6.00	90.4	11.8%	106.3
4/1/1994	6.19	87.7	17.8%	160.4
4/8/1994	6.47	87.8	22.4%	201.8
4/15/1994	6.47	87.5	22.7%	204.6
4/22/1994	6.60	88.2	24.3%	218.8
4/29/1994	6.56	87.3	24.5%	220.1
5/6/1994	6.76	84.8	30.4%	273.5
5/13/1994	6.98	85.3	33.7%	303.0
5/20/1994	6.65	87.3	26.0%	234.1

they would have ever seen the VaR for Leeson's trades because Leeson himself was responsible for recording his trades and it appears as though he was hiding his trading activity from senior management. Had they had a proper VaR system in place, complete with an IT system to capture, record and report on all trading activity, they might have been able to detect his unauthorized activity.

Orange County

This municipality suffered $1.5 billion in losses when its leveraged bets on the slope of the yield curve went sour. There is some debate though regarding what really caused the loss. Some experts reviewed the Orange County portfolio and noted that the fund still had assets worth $20 billion, a net worth of $6 billion and plenty of cash to meet its needs. They argue that the forced liquidation of the portfolio is what really caused the losses — within a year as short term rates declined the loss would have been fully recovered.

On the other side of the debate though, Jorion showed that the one year VaR would have suggested that the potential for more than $1 billion in losses was so high that the investment strategy would have been called into question much earlier. But that presupposes that Orange County would have been focused only on the potential downside in complete isolation from the potential upside from the strategy. If the fund manager was knowingly taking a large risk in order to earn high returns, as he appeared to be doing, then perhaps VaR would not have motivated any change in strategy.

This highlights an important point about VaR — for entities whose mission is to take risks, reporting the potential losses alone without also reporting the potential gains renders the information rather meaningless.

Metallgesellschaft (MGRM)

MGRM sold long term fixed price oil contracts to customers and hedged the risk using short term futures contracts. This was an example of what Stulz referred to as a speculative hedge — they used their knowledge of the oil market to speculate on the basis (price differentials in different markets) but hedged the level of oil prices. But when prices fell and the futures contracts had huge paper losses (offset for the most part by unrecognized huge paper gains on the long term contracts sold to their customers), senior management forced a liquidation of the futures contracts.

Could VaR have helped? It could have told them how large of a bet they were taking on the basis and rollover risk (replacing short term futures contracts as they expire with new short term futures contracts), but that's the business they were in and they were taking those risks intentionally and apparently with full knowledge of the size of the risk.

But again, VaR is a value risk measure. The real problem for MGRM wasn't the value of the positions (remember, the losses were in theory at least offset by the gains on the customer contracts), but rather the cash flow requirements of the positions. Had they better understood the factors that might have forced them to liquidate the positions, perhaps they would have adopted a different strategy. VaR wouldn't have told them this though.

Alternatives to VaR

If VaR would not have been of much help in avoiding those four disasters, what would have? Three alternatives are presented:

- Cash Flow Risk — Simulating future cash requirements (e.g. via a DFA model) can be a useful tool for firms concerned with cash flow variability, either for setting prudent debt levels or establishing contingent funding arrangements.

- Risk Based Capital — Selected risk managers or entities that are knowingly taking risks (e.g. Orange County and MGRM) are overly penalized by VaR because it does not take into account the potential gains. A better approach is to establish risk based capital measures and to evaluate the opportunities in relation to those measures. Of course, implementing this is much more difficult than implementing VaR because it is not clear what the appropriate *cost of capital* might be for different risks.

- Shortfall Risk — Rather than measure a single (arbitrary) point on the profit and loss distribution it may be preferable to use a measure that incorporates outcomes that are below some target return or, more generally, below some other meaningful quantity (e.g. planned quarterly profit). Anchoring on a meaningful number and then focusing on either the probability or the amount of deviations from that can produce a risk measure that is easier to interpret.

  Two specific examples of this general point are provided. One, which is referred to as the Below Target Probability (BTP), is just the probability of having a shortfall relative to the specified target. The other, referred to as Below Target Risk, measures the average value of outcomes below the target, which is closely related to the CTE and EPD risk measures.

  Note that the term "shortfall" is used in different ways by different authors. Although Culp, Miller and Neves used the term in the context of setting a more meaningful reference point, some authors use the term shortfall risk measure to refer to measures like CTE or Tail VaR, which average the value of all outcomes that are below some specified percentile (VaR point). Others use it similarly to measure only the amount by which the outcome exceeds a specified percentile. They typically argue that incorporating all of the outcomes gives a more meaningful risk measure, particularly for skewed distributions, and that they have better mathematical properties (they are *coherent*). Both are valid uses of the term.

Practice Questions

Question 1. Value at Risk had three features that were of great appeal to trading firms. What were they?

Solution. The three features of VaR discussed were:

- Consistent risk measure across different markets, business and products;
- Probability based; and
- Common time horizon.

Question 2. List the five uses of VaR discussed in the paper.

Solution. The five uses were:

- Risk Reporting — Its simplicity allowed for quick and efficient risk reporting to senior management.

- Risk Control — It also proved useful as a means to monitor and set risk levels by market, by trading group or by counterparty.

- Risk Management — VaR provided useful information for hedging risks and evaluating any particular transaction's effect on portfolio risk.

- Capital Allocation — VaR has also been useful to measure risk adjusted returns for compensation purposes or for capital allocation purposes.

- Exposure Monitoring — Non-trading firms tend to use measures like VaR more for monitoring external fund managers or assessing collateral requirements for customers.

Question 3. The authors note that VaR is not very useful for speculative hedging or for entities that knowingly take risks. Why?

Solution. VaR only focuses on the potential losses without any consideration given to the potential gains. It is therefore not a useful measure for balancing the risk-reward trade-offs.

Question 4. The authors cite four well-known derivatives disasters involving P&G, Orange County, Barings and Metallgesellschaft. What reasons do they give in each case why the use of VaR would have been unlikely to prevent these disasters?

Solution. Some of the reasons included the following:

- P&G — Had P&G used VaR for their entire swap book and monitored the additional VaR for each new transaction, they might have avoided the large bet they made on interest rates. However, VaR was never intended to be used at the transaction level and it is unlikely that a firm focused on cash flow risk management would have had the monitoring systems in place to approve swap transactions based on the VaR.

- Orange County — It is argued that the Orange County treasurer was intentionally taking a bet on interest rates in order to improve his portfolio returns. Since VaR focuses only on the downside without any mechanism to assess the risk in relation to the potential gains, this risk measure would have been unlikely to motivate any change in strategy.

- Barings — Knowing the VaR of Leeson's trading book might have prevented Barings' losses, but unfortunately Barings completely lacked any internal controls or information systems that would have passed this information to senior management. Leeson himself was responsible for monitoring his portfolio and it is probably safe to assume that he would not have reported his VaR to his superiors.

- Metallgesellschaft — Again, VaR could have provided quantitative information about the size of the rollover and basis risk, but MGRM was in the business of taking these risks and appeared to be doing it knowingly. It seems unlikely that whatever additional information contained in the VaR would have altered their strategy. More importantly, VaR is a value risk measure and what MGRM really needed was a better understanding of the cash flow risks associated with their strategy.

Question 5. What three alternative risk measures do the authors suggest besides VaR?

Solution. The three alternatives are:

- Cash Flow Risk
- Risk Based Capital
- Shortfall Risk

Question 6. Culp, Miller & Neves discuss the use of VaR by firms other than the trading firms where the VaR measures were first developed but notes that they often use it for other purposes than simply risk measurement. What do they say these firms use VaR for?

Solution. They argue that value risk managers benefit more from using VaR as an exposure monitoring tool, such as for policing external money managers or setting collateral requirements for customers.

Question 7. Culp, Miller & Neves discuss two approaches to risk management. What are these two approaches?

Solution. Firms can either be value risk managers, concerned primarily with total firm value at any point in time, or cash flow risk managers, concerned primarily with cash flow variability.

Question 8. Culp, Miller & Neves discuss certain characteristics shared by value risk managers. What characteristics make VaR a particularly appropriate risk measure for them?

Solution. The three characteristics mentioned are:

- They are interested in the value of their positions over short periods of time. The short time horizons imply that VaR measurement can be accomplished reliably with minimal concern about changing portfolio composition over the risk horizon.
- They need to limit and control their exposures.
- Many value risk managers are involved in agency transactions where they act primarily as intermediaries rather than take their own proprietary positions (i.e. they act on behalf of others rather than for their own account or with their own capital).

Question 9. What do Culp, Miller & Neves consider to be the advantages and disadvantages of Stulz's suggestion to allocate capital to risky activities and measure abnormal returns based on the cost of new equity needed to finance the risky activity.

Solution. On the positive side, Stulz's method does not penalize selective risk managers taking advantage of a comparative/informational advantage as VaR does. But they doubt the method could ever be implemented, given the inherent difficulty in estimating the cost of equity capital. They also take issue with the reliance on equity returns only, since projects should be evaluated based on their risks and not on their financing source.

Hull Chapter 23: Credit Value at Risk

Default Correlation

It is important to understand the degree to which two companies' default probabilities are correlated with each other. There are two general approaches that are used to model default correlations:

- *Structural* models, such as the Merton model, correlate the stochastic processes the assets of the firms follow.

- *Reduced Form* models assume that the hazard rates (the conditional default probabilities) for different companies follow a stochastic process and are correlated with macroeconomic variables. These can be used with either real world or risk neutral probabilities.

Reduced Form Models

Because reduced form models result in correlated probabilities, the actual default experience generated may be uncorrelated. In other words, the probabilities of default for two bonds might both be high, but the actual default event does not necessarily have to occur at the same time.

One example of this is the Gaussian copula model for the time to default. A *copula* is simply a multivariate distribution of two or more random variables which are each between 0 and 1. It can be used to describe the degree to which two or more probabilities are dependent on each other.

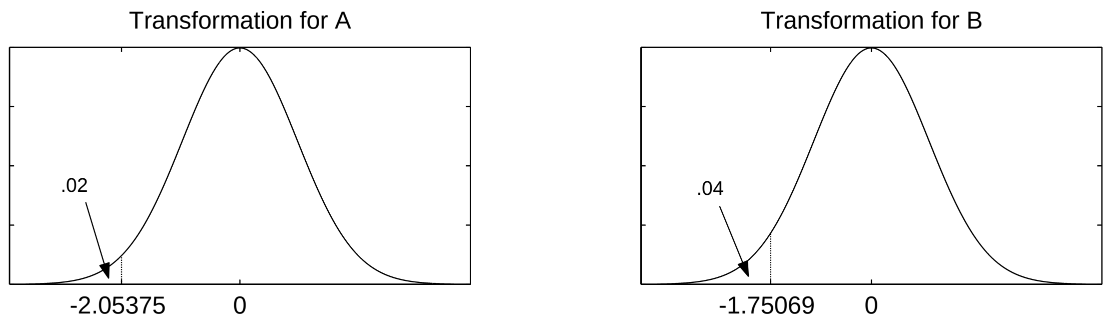
Gaussian Copula

Hull discusses the Gaussian copula, which is based on the multivariate normal distribution. Suppose t_A and t_B are the *times to default* for companies A and B. We can define $x = N^{-1}[Q(t)]$ for each company, where $Q(t)$ is the cumulative probability of default through time t and $N^{-1}[\cdot]$ is the inverse of the standard normal distribution function.

Since $Q(t)$ is a probability between 0 and 1, the inverse normal function will transform this into a normally distributed random variable. The joint probability of A and B defaulting can be generated from a multivariate normal distribution by calculating the probability of observing these transformed variables x_A and x_B from a joint standard normal distribution with a correlation coefficient of ρ.

For example, assume that the probability of A defaulting by the end of year 1 is 2% and the probability of B defaulting by the end of year 1 is 4%. We want the probability that they will both default before year 1, assuming a correlation coefficient of 0.5.

FIGURE 1. Demonstration of Copula Method



To do this, we first transform the probabilities into normally distributed random variables, x_A and x_B. We set x_A to the value of a standard normal random variable that will occur with a probability of 2%, which is -2.05 and x_B to the value of a standard normal random variable that will occur with probability of 4%, which is -1.75. These transformations are shown in Figure 1.

Then, to determine the joint probability, we just find the joint probability that a multivariate normal distribution with correlation coefficient of 0.50 will take on the values $x_A < -2.05$ and $x_B < -1.75$. In this case the probability is 0.0054. That is, if the time to default for the two companies follows a Gaussian copula with a correlation of 0.5, then with a 2% probability of A defaulting and a 4% probability of B defaulting the probability that they will both default is only 0.0054. Had their times to default been independent the joint probability would have only been $(2\%)(4\%) = 0.0008$.

The above explanation showed how to calculate the joint probabilities when the two individual marginal distributions were not necessarily easy to combine into a multivariate distribution. But usually the Gaussian copula will be implemented in *reverse*. Often you will use it to simulate correlated normal random variables, the x_A and x_B values referred to above, and then transform those to the default probabilities using the distributions for the time to default.

One challenge in using the Gaussian copula is to specify the correlation coefficients to use. When there are multiple variables, this will involve a correlation matrix with numerous inputs. To make this a bit easier, the book suggests using the assumption that a one-factor model drives the correlations. This eliminates the need to estimate a correlation coefficient for each pair of variables because the coefficients in the factor model determine the correlations of any pair.

One-Factor Model

To avoid having to specify a separate correlation for each pair of bonds in a portfolio, it is helpful to assume a factor model with a single common factor that drives correlation. Each company will have a different sensitivity to this factor (we denote the sensitivity as a_i below, but this is conceptually the same as what we called ”beta” in the BKM discussion of factor models). The factor sensitivities of any two companies will determine the correlation between them.

The factor model says that whether or not a company defaults is determined by a transformed variable x_i (think of it as the default threshold in the Merton model) which can be written in terms of a common factor F and a company-specific random term Z, as follows:

$$x_i = a_i F + \sqrt{1 - a_i^2} Z_i$$

Under this assumption, the correlation between x_i and x_j is $\rho = a_i a_j$.

When the following is true,

i. the common factor F follows a standard normal distribution,

ii. the default distributions are assumed to all be the same for all companies and

iii. the correlations are assumed to all equal $\rho = a_i a_j$, which also means that $a_i = \sqrt{\rho}$

then the probability of default for each company at time T can be written as follows for a given value of F:

$$\begin{aligned} Q(T|F) &= Pr[a_i F + \sqrt{1 - a_i^2} Z_i < \Phi^{-1}(Q(T))] \\ &= Pr\left[Z_i < \frac{\Phi^{-1}(Q(T)) - a_i F]}{\sqrt{1 - a_i^2}} \right] \\ &= \Phi\left[\frac{\Phi^{-1}(Q(T)) - \sqrt{\rho} F}{\sqrt{1 - \rho}} \right] \end{aligned}$$

Notice in the last step above I used the fact that we assumed Z was normally distributed and also substituted $a_i = \sqrt{\rho}$ to write the formula in terms of ρ rather than a_i.

Credit Value at Risk

The N-year, X% Credit VaR attempts to identify the dollar amount of credit related (default) losses that will not be exceeded with X% probability over an N-year period. If you had a probability distribution for the credit-related losses over the specified time period, the Credit VaR would be the $1 - Xth$ percentile.

Before getting into the details, note a few points:

- Credit VaR attempts to determine a dollar amount that our credit losses will not exceed with some high probability.

- Unlike traditional market VaR where the time horizon is commonly assumed to be relatively short, such as 10 days or less, the time horizon for Credit VaR is usually a year or more.

- Here, we are trying to determine the dollar amount of the loss, not its value, and therefore we use real world probabilities rather than risk neutral probabilities.

- Two methods are discussed, one of which uses the Gaussian copula default model with a common factor driving correlations, as described above, and one that uses credit rating transition probabilities.

Gaussian Copula Model for Credit VaR

In the Gaussian copula model and under some simplifying assumptions, Hull showed that:

$$Q(T|F) = \Phi\left[\frac{\Phi^{-1}[Q(T)] - \sqrt{\rho}F}{\sqrt{1-\rho}}\right]$$

This quantity represents the probability of default for any specific company. If the number of companies in the portfolio is infinitely large, then the **percentage of defaults in the portfolio** by time T converges to this same value.

To use this model, we don't actually need to know the value for F. If we assume it is a standard normal random variable, then we know that its value will be greater than $-\Phi^{-1}(X\%)$ with a probability of $X\%$. Therefore, the percentage of defaults in the portfolio will be less than the value $V(X,T)$ $X\%$ of the time, where:

$$V(X,T) = \Phi\left[\frac{\Phi^{-1}[Q(T)] + \sqrt{\rho}\Phi^{-1}(X)}{\sqrt{1-\rho}}\right]$$

This formula looks very similar to the previous formula for $Q(T|F)$ but notice the sign change in the numerator!

To calculate the Credit VaR, we simply need to also reflect the value of the portfolio and the average recovery rate in the event of default.

For instance, assume the portfolio is worth \$100 million, the one-year probability of default is 2% (for every asset in the portfolio) and the recovery rate is 60% in the event of default. If the Gaussian copula correlation coefficient is $\rho = 0.10$, then we can be 99.9% certain that the percentage of defaults will be less than the following quantity:

$$\begin{aligned}
V(X,T) &= \Phi\left[\frac{\Phi^{-1}[Q(T)] + \sqrt{\rho}\Phi^{-1}(X)}{\sqrt{1-\rho}}\right] \\
&= \Phi\left[\frac{\Phi^{-1}[.02] + \sqrt{.1}\Phi^{-1}(99.9\%)}{\sqrt{1-.1}}\right] \\
&= \Phi(-1.1348) \\
&= .128
\end{aligned}$$

Therefore, the 1-year, 99.9% credit VaR is equal to:

$$(\$100 \text{ million})(.128)(40\%) = 5.13 \text{ million}$$

CreditMetrics Model for Credit VaR

The second method (used in the CreditMetrics model) also produces measures of credit risk similar to the Value at Risk measure we used earlier. This approach simulates credit rating changes for each counterparty using a credit transition matrix. The credit transtion matrix depicts the probability of any change in the rating, not just the probability of default.

As an example, consider the historical credit-rating transition data for AA-rated bonds in Table 1.

TABLE 1. Credit Rating Transition

	Rating at Year-End							
Initial Rating	AAA	AA	A	BBB	BB	B	CCC	D
AA	0.66	91.72	6.94	0.49	0.06	0.09	0.02	0.01

This table indicates that a AA-rated bond has a 91.72% probability of still being AA-rated at year end, a .49% probability of being BBB-rated, a .01% probability of defaulting, etc.

Using a matrix like this for all other initial ratings, we could easily simulate the results for an entire portfolio bonds over one year or more. A Gaussian copula can be used to correlate the credit rating transitions for different counterparties and then estimates of forward yield curves for each rating could be used to revalue the bonds at their simulated new ratings. The change in value over the time period can then be computed for each simulated set of credit ratings. The distribution of the change in value can be used to calculate the Credit VaR using the same approach as we used for market VaR.

Notice a difference between the Gaussian copula model for Credit VaR and the CreditMetrics model for Credit VaR. In the copula model, only defaults are modeled. In the CreditMetrics model, both defaults and credit downgrades are reflected.

Practice Questions

Question 1. Suppose you have *time to default* distributions for Aa-rated and Baa-rated bonds as shown in the table below, based on historical default probabilities published by Moody's.

TABLE 2. Average Cumulative Default Rates (%), 1970-2003

	1	2	3	4	5	10	20
Aa	0.02	0.03	0.06	0.15	0.24	0.68	2.70
Baa	0.20	0.57	1.03	1.62	2.16	5.10	12.59

Both bonds have a 20-year maturity. Assuming that you randomly generate two independent uniform random numbers with values $U_1 = .242$ and $U_2 = .018$, what would be the simulated time to default for each of these bonds? In other words, determine whether there is a default and also determine when it occurs.

Solution. For the Aa-rated bond, there is a 2.70% probability of default during its 20-year life. If we were to randomly generate a uniform random number less than or equal to 2.70%, then we would say that there was a default. In this case, the number $U1 = .242$ and so we would conclude that there is not a default.

For the Baa-rated bond, the simulated value is equal to $U_2 = .018$. From the table, we see that there is a 1.62% probability of default by year 4 and a 2.16% probability of default by year 5. The simulated value of .018 therefore suggests that a default occurs between years 4 and 5.

Question 2. Suppose you wanted to use a Gaussian (normal) copula to generate two dependent uniform random numbers, using a correlation coefficient of $\rho = 0.60$. You are given the following sequence of independent uniform random numbers: 0.23, 0.57. Use these with the Choleski decomposition to generate the dependent uniform random numbers.

Solution. The key to this question is to notice that it is very easy to generate correlated normal random variables, but harder to generate correlated (or dependent) uniform random numbers. But you accomplish this using an intermediate step.

You start with two independent uniform random numbers, transform those into independent normal random variables, use these to generate another normal random variable that is correlated with the first (using a Choleski decomposition), and then transform these correlated normal random variables back into dependent uniform random variables.

The first step is to create two independent normal random variables using the values 0.23 and 0.57. These are created by inverting the normal distribution function, so:

$$x_1 = \Phi^{-1}(0.23) = -0.739$$

$$x_2 = \Phi^{-1}(0.57) = 0.176$$

Now, use those two independent normal random variables to create a third normal variable that is correlated with the first one. The Choleski decomposition formula is:

$$\begin{aligned} x_3 &= \rho x_1 + \sqrt{1-\rho^2}x_2 \\ &= .60(-0.739) + \sqrt{1-.60^2}(0.176) \\ &= -0.3026 \end{aligned}$$

Now, use the normal distribution function to transform these back into uniform random numbers, to get:

$$U_1 = \Phi(-0.739) = 0.23$$

$$U_2 = \Phi(-0.302) = 0.38$$

Notice that the first uniform random number is simply the first value we started with. The second one however is lower than one we independently generated, since it was pulled closer to the first value. The correlation coefficient determines the strength of that pull. If the correlation had been 1.0, then the second value would have also been 0.23. If it had been 0.0, then it would have remained at 0.57.

Notice too that the calculation does NOT apply the Choleski decomposition to the uniform random numbers.

Question 3. Suppose you have two Ba-rated bonds with 20 years to maturity. Each bond has a time to default probability distribution as shown in the following table:

TABLE 3. Average Cumulative Default Rates (%), 1970-2003

	1	2	3	4	5	10	20
Ba	1.26	3.48	6.00	8.59	11.17	21.01	38.56

Assuming a normal copula with correlation of 0.60, use the results from the previous question to simulate correlated times to default for each bond.

Solution. The previous question used a normal copula to generate dependent uniform random numbers equal to 0.23 and 0.38. Using these, one bond would have a simulated time to default of approximately 11.1 years (using linear interpolation) and the second bond would have a time to default of approximately 19.7 years.

Question 4. Hull discusses the fact that to use the Gaussian copula to simulate a large portfolio of bonds, you would need to begin with an entire correlation matrix (and use a more elaborate version of the Choleski decomposition).

As a simplification, he suggests using a single-factor model such that the correlation among any two pairs of bonds is determined based on the sensitivity of each bond to this single factor. In the case where the default probability is the same for all bonds, the correlation between any two bonds is the same, and the number of bonds in the portfolio is infinitely large, then X% of the time the percentage of bonds that default will be less than the following quantity:

$$V(X, T) = \Phi\left[\frac{\Phi^{-1}[Q(T)] + \sqrt{\rho}\Phi^{-1}(X)}{\sqrt{1-\rho}}\right]$$

Use this formula to estimate the 95%, one-year Credit VaR for a $250 million portfolio of bonds, if each bond has a 1% one-year default probability, the recovery rate is 70% and the Gaussian copula correlation coefficient is $\rho = 0.20$.

Solution. Using the formula given, we can be 95% certain that the percentage of defaults will be less than:

$$\begin{aligned} V(X, T) &= \Phi\left[\frac{\Phi^{-1}[Q(T)] + \sqrt{\rho}\Phi^{-1}(X)}{\sqrt{1-\rho}}\right] \\ &= \Phi\left[\frac{\Phi^{-1}[.01] + \sqrt{.2}\Phi^{-1}(95\%)}{\sqrt{1-.2}}\right] \\ &= \Phi(-1.7785) \\ &= 0.0377 \end{aligned}$$

Therefore, the 1-year, 95% Credit VaR is equal to:

$$(\$250 \text{ million})(.0377)(30\%) = \$2.82 \text{ million}$$

Note: The solution above did not use rounding for the inverse of the normal distribution. If you had used $\Phi^{-1}(.01) = -2.33$ *and* $\Phi^{-1}(.95) = 1.645$*, your final answer would have been closer to \$2.8 million.*

Question 5. The following is a one-year credit rating transition probability table for a company currently rated AA. Assume that we wanted to simulate the credit rating for this company at the end of 1 year and we randomly generated a value of 1.61 from a standard normal distribution function. What would be the rating for this company at the end of 1 year?

TABLE 4. Credit Rating Transition

	Rating at Year-End							
Initial Rating	AAA	AA	A	BBB	BB	B	CCC	D
AA	0.66	91.72	6.94	0.49	0.06	0.09	0.02	0.01

Solution. First, we can extend the table above to show the cumulative probabilities that the bond rating is equal to or better than each possible rating:

TABLE 5. Cumulative Transition Probabilities

Initial Rating	AAA	AA	A	BBB	BB	B	CCC	D
AA	0.66	91.72	6.94	0.49	0.06	0.09	0.02	0.01
Cumulative	0.66	92.38	99.32	99.81	99.87	99.96	99.98	100.00

To clarify, note that the table indicates a .66% probability of a AA bond becoming AAA by the end of the year, a 91.72% probability of staying AA, a 6.94% probability of becoming A, etc. The values in the second row indicate the cumulative probabilities of migrating to a rating at least as high as the rating in each column, so there's a 99.32% probability of being rated at least A, for example.

The simulated standard normal random variable of 1.61 has a cumulative probability of $\Phi(1.61) = .9463$. From the transition probability table, there is a 92.38% probability that the AA bond will be rated AA or better at the end of the year and a 99.32% probability of being rated at least A. The simulated probability of .9463 therefore suggests the year-end rating will be worse than AA but not worse than A. Without any intermediate rating classes, this suggests an A rating.

Question 6. Suppose you wanted to simulate another year-end credit rating for a different AA-rated bond than the one in the previous question. Assume that the transition probabilities for the two bonds had a correlation of .65 and use a Gaussian (normal) copula. Also assume that a new simulated standard normal random variable was independently generated and yielded a value of .24. Determine the rating for this second bond at the end of one year.

Solution. The previous question used a simulated normal random variable of 1.61. We can generate a second standard normal random variable that has a correlation of .65 with the first random variable using a Choleski decomposition.

The formula to use in this case is:

$$N_2 = .65(1.61) + \sqrt{1 - .65^2}(.24) = 1.229$$

Notice what this has done so far. We have simulated two standard normal random variables that have a correlation of 0.65 with each other. The two values are $N_1 = 1.61$ and $N_2 = 1.229$. The second value was based on the first value, an independent standard normal and the correlation coefficient.

We can now transform these correlated normal random variables into correlated (technically we should say *dependent*) uniform random variables using the standard normal distribution to get:

$$U_1 = \Phi(1.61) = .9463$$

$$U_2 = \Phi(1.229) = .89$$

The credit ratings are then determined as in the previous question. The first bond will be rated A (see previous question) and the second bond will be rated AA.

Butsic: Solvency Measurement for Risk Based Capital Applications

This paper was written to help establish Risk Based Capital requirements for P&C insurance companies, but the primary reason it is on the syllabus is because of its coverage of the Expected Policyholder Deficit (EPD) as a means to measure the amount of capital required to ensure an acceptably low level of insolvency risk. The part to focus on the most is the measurement of the EPD and its important characteristics.

Risk Based Capital

In an efficient market, where information is costlessly obtained and solvency risk is transparent, the risk of an insurer not meeting its obligations to policyholders would be reflected in the cost of insurance (as a reduction in premium as this risk increased). However, given the complexity of analyzing the solvency risk of P&C insurers and the limitations on the typical policyholder (the consumers), insurer solvency is considered to be a regulatory issue and risk based capital requirements provide a minimum level of solvency protection to the various parties (insured, claimants, capital providers).

The criteria for an effective risk based capital method include:

- It should be the same for all classes of insureds (personal vs. commercial), all types of insurers (primary insurance vs. reinsurance) and all types of claimants.
- It should be objectively measured — based on insurer's financial data and a mathematical formula.
- It should discriminate between quantifiable measures of risk.

Expected Policyholder Deficit (EPD)

The measure Butsic recommends is the EPD, which is defined as the expected value of the difference between the insurer's obligation to pay the claimant and the actual amount paid. In other words, it is the expected shortfall in payments that result from inadequate insurer capital.

Mathematically for a discrete loss distribution:

$$\text{Expected Policyholder Deficit} = D_L = \sum_{x>A} (x - A)p(x)$$

where D_L is the EPD, x is the size of loss, A is the value of the assets (which are assumed constant at this point) and $p(x)$ is the probability density of x.

The calculation of the EPD and the EPD ratio proceeds as shown in Table 1 on the next page for the simple case with fixed assets (non-stochastic) and loss amounts can take on one of only three values. Note that in this calculation, the Claim Payment is the actual amount paid given

TABLE 1. EPD Calculation: Fixed Assets

Scenario	Asset	Loss	Capital	Probability	Claim Payment	Deficit
1	13,000	6,900		0.2	6,900	0
2	13,000	10,000		0.6	10,000	0
3	13,000	13,100		0.2	13,000	100
Expected Value	13,000	10,000	3,000		9,980	20
					EPD	20
					Expected Loss	10,000
					EPD Ratio	0.002

the limitation of the asset values. The deficit is then defined as the amount of loss not paid, or max(Loss − Claim Payment, 0).

Also note that in Butsic's tables, the amount shown in the "Capital" column is the expected assets less the expected losses.

EPD Ratio

The EPD can apply equally to the measurement of asset risk and liability risk. To adjust the scale to reflect different risk element sizes, the ratio of the EPD to the expected loss is used as the basic measure of policyholder security. Mathematically, the EPD ratio is given as:

$$\text{EPD Ratio} = d_L = D_L/L$$

where D_L is the EPD and L is the expected loss.

Probability of Ruin

Traditional risk measures for insurers tended to revolve around the probability of ruin, or the probability that the insurer would not be able to fully satisfy its obligation. For the insurance company managers who care primarily whether they will be employed or not, this might be a reasonable measure of risk. However, for the claimant, what is of primary interest is the *amount* of the shortfall. As Butsic says, it is not sufficient merely to consider the probability of ruin — its severity must also be appreciated. While certain strategies may cause the *probability of a shortfall* to decrease, this may not mean that our risk is reduced if at the same time the *severity of the shortfall*, when it occurs, is increased.

EPD with Asset Risk

The example above assumed that the asset values were constant and only the loss amounts were stochastic. If instead we assume that the loss amounts are constant and the asset values are stochastic, then we would proceed in essentially the same fashion, as shown in Table 2 on the facing page.

TABLE 2. EPD Calculation: Fixed Losses

Scenario	Asset	Loss	Capital	Probability	Claim Payment	Deficit
1	12,000	5,000		0.1	5,000	0
2	6,000	5,000		0.8	5,000	0
3	3,000	5,000		0.1	3,000	2,000
Expected Value	6,300	5,000	1,300		4,800	200
					EPD	200
					Expected Loss	5,000
					EPD Ratio	0.04

Note that even though we are varying the asset value in this example, the EPD Ratio is still calculated as a ratio to the expected loss.

Setting Capital Requirements

Suppose that we wanted to set the EPD ratio to 5% for both of the examples above. To do this, we merely increase or decrease the asset amounts so that the resulting EPD ratio is 5%. In the cases shown above, the EPD ratios were lower than 5% so in both cases we are going to reduce the starting assets.

The calculation can be done for the two examples above in a spreadsheet using the Solver or Goal Seek commands. Note that it seems that you could do this algebraically, but there's a subtle complication. In Butsic's Table 3 on page 662 where he shows the results of this calculation, he greatly simplified his own calculations because he noticed (assumed) that Scenarios 1 and 2 never caused a deficit and therefore he was able to ignore them for his calculation. But because you are possibly lowering the asset values to meet the target EPD ratio, it is not generally safe to do this — you may not have a deficit in those scenarios to begin with, but with a low enough new asset value you might create one. Therefore, I will show the whole table here.

Fixed Assets, Stochastic Liabilities

In Table 3 on the next page I show the assets needed to produce a 5% EPD ratio. Once that amount is known, subtracting the expected loss (reflecting the current reserves) indicates the amount of capital (surplus) needed.

Stochastic Assets, Fixed Liabilities

In this case, the calculations are a bit tricky and Butsic could have been clearer about what he did. In his examples where the assets were stochastic, it is their *ending* values that are stochastic. Their beginning values just happened to be assumed to equal the expected value of the ending asset values, or \$6,300. So in other words, he assumed that in Scenario 1 the beginning asset value is \$6,300 and that the ending value was \$12,000, or \$12,000/\$6,300 = 190.5% of the beginning value. For Scenarios 2 and 3, he was similarly assuming that the ending

TABLE 3. Capital Needed for 5% EPD Ratio: Fixed Assets

Scenario	Asset	Loss	Capital	Probability	Claim Payment	Deficit
1	10,600	6,900		0.2	6,900	0
2	10,600	10,000		0.6	10,000	0
3	10,600	13,100		0.2	10,600	2,500
Expected Value	10,600	10,000	600		9,500	500
					EPD	500
					Expected Loss	10,000
					EPD Ratio	0.05

assets were 95.2% and 47.6% of the beginning values. So when he adjusted the beginning asset values, he assumed that the return distribution did not change, i.e. the returns and probabilities stay the same.

Table 4 shows this a bit more clearly with some additional columns. Note that the example

TABLE 4. Capital Needed for 5% EPD Ratio: Fixed Losses

Scenario	Beginning Asset	Return	Ending Asset	Loss	Capital	Probability	Claim Payment	Deficit
1	5,250	190.50%	10,000	5,000		0.1	5,000	0
2	5,250	95.20%	5,000	5,000		0.8	5,000	0
3	5,250	47.60%	2,500	5,000		0.1	2,500	2,500
Expected Value		100.00%	5,250	5,000	250		4,750	250
							EPD	250
							Expected Loss	5,000
							EPD Ratio	0.05

above required us to decrease the amount of invested assets and we assumed that the distribution of the asset returns stayed the same. In cases where we need to increase the invested assets, you would usually do the same thing. However, if you were to assume that you were going to invest the additional assets in risk free securities rather than in assets with the same distribution as the current assets, then you would not need to worry about the ratio of ending to starting assets. The Ending Risky Asset amounts would stay the same as in the table and you would merely add a constant amount of assets to each of the scenarios. See the questions below for an example of this.

Stochastic Assets and Stochastic Liabilities

When the assets and liabilities are both stochastic, then the calculation of the EPD proceeds in precisely the same fashion, though now you need to use the *joint distribution* of the policyholder deficit $\max(0; L - A)$.

In the case of discrete distributions, this is trivial because you simply need to build a table with the possible asset values in the columns, the possible liability values in the rows and each cell of the table then contains the value of the deficit and its associated probability. In the simple case with independent asset and loss values, the probability for each combination is simply the product of the corresponding asset and liability probabilities. See the questions below for an example of this.

When the distributions are continuous, the math is much more involved, though the logic is identical.

Continuous Distributions

The extensions of the EPD formulas for assets and liabilities are rather trivial if the distributions are assumed to be continuous and should be familiar to you given the parallel to the expected excess insurance losses covered on earlier exams.

The formulas are:

$$\text{Risky Liabilities, Fixed Assets:} \qquad D_L = \int_A^\infty (x - A)p(x)\,dx$$

$$\text{Risky Assets, Fixed Liabilities:} \qquad D_A = \int_0^L (L - y)p(y)\,dy$$

Analytic Formulas for EPD Ratios (Normal Distribution)

We will assume that the assets or liabilities are normally distributed and define the following variables:

- c = ratio of capital to expected losses, such that $A = L + C = (1 + c)L$
- $c_A = C/A$ = capital to expected asset ratio
- k = coefficient of variation (ratio of standard deviation to mean) for the liabilities
- k_A = coefficient of variation for the assets
- $\phi(\cdot)$ is the standard normal density function

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

- $\Phi(\cdot)$ is the standard normal distribution function

The Appendix to the Butsic paper shows that under these assumptions we have the following simplified formulas for the EPD Ratios:

$$d_L = \frac{D_L}{L} = k\phi\left(-\frac{c}{k}\right) - c\Phi\left(-\frac{c}{k}\right)$$

$$d_A = \frac{D_A}{L} = \frac{1}{1-c_A}\left[k_A\phi\left(-\frac{c_A}{k_A}\right) - c_A\Phi\left(-\frac{c_A}{k_A}\right)\right]$$

Notice that to establish the amount of capital required for a given EPD Ratio, we need to simply solve the above formulas for c. As shown in Figure 1 of the Butsic paper, the Capital/Loss ratio corresponding to a given EPD Ratio is approximately a linear function of the coefficient of variation, k.

Relationship to Ruin Probabilities

Recall the probability of ruin if assets are constant and liabilities are risky is:

$$\text{Prob}[C < 0] = \text{Prob}\left[X < \left(\frac{0 - E(C)}{\text{StdDev}(C)}\right)\right]$$

where X is a normal random variable.

But note that $C = A - L$ and therefore since only L is random, $E(C) = A - E(L)$ and $\text{StdDev}(C) = \sigma_L$. This gives us the following ruin probability:

$$\begin{aligned}\text{Ruin Prob} &= \Phi\left(-\frac{C}{\sigma_L}\right)\\ &= \Phi\left(-\frac{C/L}{k}\right)\\ &= \Phi\left(-\frac{c}{k}\right)\end{aligned}$$

Note from the formula for the ruin probability that there will be a different ruin probability for different values for c and k. But the EPD ratios are constant across a wide range of c and k values, suggesting that there is no single ruin probability corresponding to a given EPD ratio.

EPD Ratios Under Lognormal Distribution

It might be more realistic to assume, particularly for losses, that the risk element is lognormally distributed rather than normally distributed. In this case, the Appendix uses a parallel to the Black-Scholes option pricing formula to show the formulas for the EPD ratios as functions of the capital to loss ratios and the coefficient of variation.

The formulas given are:

$$d_L = \Phi(a) - (1+c)\Phi(a-k)$$

$$d_A = \Phi(b) - \frac{\Phi(b-k_A)}{1-c_A}$$

where,

$$a = \frac{k}{2} - \frac{\ln(1+c)}{k}$$
$$b = \frac{k_A}{2} + \frac{\ln(1-c_A)}{k_A}$$

To understand the derivation of these formulas, we'll use Butsic's option pricing parallel. We'll look at the cases with risky liabilities and with risky assets separately.

Risky Liabilities

In this case, we have fixed assets equal to A and liabilities with an expected value of L. To determine the EPD we want to know the expected value of the amount by which the liabilities exceed the assets and recall that so far Butsic has assumed that interest rates are zero. The expected value is found by solving the equation:

$$D_L = \int_A^{\infty} (x-A)p(x)\,dx$$

Recall that a call option with an exercise price of K on a stock with current value of S and a lognormally distributed value at expiration S_T has a value today that can also be found in a similar fashion. Using the Risk Neutral Valuation approach, we know that the value of the option is the discounted expected value of the payoff, where the expectation is taken with respect to a risk-neutral probability distribution.

The formula for this is:

$$\text{Call} = e^{-rT}\int_K^{\infty} (S-K)p(S)\,dS$$

Note that in the case where the time value of money is zero, the e^{-rT} term can be ignored and we have a formula for the call option that looks just like the formula for the EPD, but with the exercise price K replaced by A and the stock price S replaced by L. Therefore, we can use the solution to this integral that we know from the Black-Scholes formula and just replace K with A, S with L, etc.

Let's first make sure we understand the logic of why the EPD is like a call option and then we'll deal with the math in the appendix to these notes.

The EPD represents the value of the shortfall between the loss amount and the assets of the insurance company. This value represents the cost to the policyholder of a (partial) default by the insurance company, in which case they would get the assets of the firm instead of their full loss (claim) payment.

In the case of fixed assets and risky liabilities, the insurance company has the right to pay the fixed amount A to settle an obligation worth L when L turns out to be worth more than A. This is just like a call option on the liabilities with a strike price equal to A.

This sometimes confuses people because the call option involves buying an asset instead of settling a liability, but these are essentially the same thing.

Risky Assets

The situation with risky assets and fixed liabilities is very similar, though the option pricing analogy is a bit different. Here, the assets are the random variable and you should think of the situation as follows. In the event the insurer's assets are insufficient to pay the loss amount, the insurer can give all of the assets to the policyholder in exchange for extinguishing its liability, L. This is as if the insurer is *selling* the assets to the policyholder for a fixed amount, L. In other words, this is like a put on the assets with a strike price equal to L.

To derive the value of this put, we note that the Black-Scholes formula can give us the value of the comparable call on the Assets with a strike price of L, and then from Put-Call Parity we know that:

$$\text{Put} = \text{Call} + PV(L) - A = \text{Call} + L - A$$

in the case where $r = 0$. The text for some reason makes this very confusing, but the derivation is trivial. See Question 5 for details.

Accounting Bias

Butsic recommends that market value accounting should be used to set capital standards and that biases from items like the discount in loss reserves should be removed.

Time Horizon

For there to be risk, there must be the passage of time, but the solvency standard must be measured using balance sheet values that reflect a point in time. And since the degree of risk depends on the time interval between evaluation points, and the spread of asset and liability values increase with time (when measured in dollars), it is important to use a common time horizon.

Capital Adjustment Through Time

The concept of an EPD Ratio standard can be used to determine the required capital adjustments through time by simply assuming that capital can be added or removed at future dates as if the insurer were being liquidated and that the reserves at the time of liquidation equaled their expected value at maturity.

Insurer as a Going Concern

Butsic further adjusts his approach to reflect the risk of new business written between evaluation points. He doesn't fully reflect the value of future new business that can be written after the hypothetical liquidation date because this would be impractical.

Correlation & Independence of Risk Elements

Up until this point, Butsic has assumed that either the losses are risky or the assets are risky, but not both, and that there was only one asset and one liability distribution. In that discussion, he argued for each risk element to be calculated separately, but unless they are perfectly correlated, the combined capital required is not the sum of the individual capital requirements.

To deal with multiple risky lines of business, multiple risky asset classes or both, Butsic does two things. First, he treats risk elements *on opposite sides of the balance sheet* as having opposite signs in terms of their correlation. Therefore, if two risk elements are on opposite sides of the balance sheet and are perfectly correlated, their correlation is assumed to be *negative* 1.0. The variance then for two liabilities, two assets or an asset and a liability are given as:

Two Assets:	$\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$
Two Liabilities:	$\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$
Asset and Liability:	$\sigma^2 = \sigma_A^2 + \sigma_L^2 - 2\rho\sigma_A\sigma_L$

IMPORTANT NOTE

Butsic is unclear about exactly when and how he adjusts the sign of the correlation coefficient.

The above formulas suggest that he measures correlation using the actual data without adjustment and then multiplies ρ by negative 1. This can be seen more clearly using his numerical example shown in Butsic's Table 13.

FIGURE 1. Butsic Table 13

Risk-Based Capital (RBC) Calculation Using
the Square Root Rule: Input Assumptions

Risk Element	*Amount*	*Capital Ratio*	*RBC*
Stocks	200	0.20	40
Bonds	1,000	0.05	50
Affiliates	100	0.20	20
Loss Reserve	800	0.40	320
Property UPR	100	0.20	20
Total			450

Correlated Risk Elements		*Correlation Coefficient*
Stocks	Bonds	0.2
Stocks	Affiliates	1.0
Bonds	Affiliates	0.2
Bonds	Loss Reserve	0.3
Affiliates	Loss Reserve	-1.0

In that table, he shows the correlation coefficient for Bonds and Loss Reserves as $\rho = .3$ and the correlation coefficient for Affiliates and Loss Reserves as $\rho = -1.0$. However, when he calculates the capital requirement of \$337, he *reverses the signs* and uses correlations of $\rho = -.3$ and $\rho = 1.0$.

Butsic further confuses the issue about the sign of the correlation coefficient for Assets and Liabilities by giving examples where the risk elements are "fully correlated", in which case he uses $\rho = +1.0$ for all risk elements regardless of which side of the balance sheet the items are on.

Secondly, he notes that the required capital under the EPD Ratio standard is approximately proportional to the standard deviation of the risk element. Therefore, he can apply his *square root rule* which sets the required capital equal to the square root of the sum of the squares of the required capital for each risk element on its own plus twice the correlation coefficient times the separate capital amounts.

This can also be extended to more than two risk elements and the formula is simply:

$$C = \sqrt{\sum C_i^2 + \sum_i \sum_{j \neq i} \rho_{ij} C_i C_j}$$

where C_i is the required capital for each risk element and the correlation coefficients, ρ, are adjusted so that if the risk elements are on opposite sides of the balance sheet their correlations are multiplied by -1.0 (i.e. the sign of the measured correlation is reversed).

Note that the formula in the text omits the double summation symbol on the covariance terms, but the calculations and the text do reflect the fact that each cross-product term appears twice.

Practice Questions

Question 1. Calculate the EPD and the EPD ratio for an insurer with fixed assets of \$13,000 and losses of either \$2,000; \$10,000; or \$18,000 with probabilities .2, .6 and .2.

Solution. The calculation proceeds just as in the notes:

TABLE 5. EPD Calculation

Scenario	Asset	Loss	Capital	Probability	Claim Payment	Deficit
1	13,000	2,000		0.2	2,000	0
2	13,000	10,000		0.6	10,000	0
3	13,000	18,000		0.2	13,000	5,000
Expected Value	13,000	10,000	3,000		9,000	1,000
					EPD	1,000
					Expected Loss	10,000
					EPD Ratio	0.10

Question 2. In the notes we used an example of an insurer with starting assets of \$6,300 and an ending asset distribution of either \$12,000, \$6,000 or \$3,000 with probabilities .1, .8 and .1. The losses were assumed to be constant at \$5,000. Show that if we lower the starting asset value to \$4,941 and assume that the asset return distribution is unchanged, the EPD ratio is 10%.

Solution. Here, I wanted to make this question relatively simple, so I gave you the answer and just asked you to confirm it. Following the same basic calculation as before but being careful with the beginning and ending asset values, we have:

TABLE 6. EPD Calculation

Scenario	Beginning Asset	Return	Ending Asset	Loss	Capital	Probability	Claim Payment	Deficit
1	4,941	190.50%	9,412	5,000		0.1	5,000	0
2	4,941	95.20%	4,706	5,000		0.8	4,706	294
3	4,941	47.60%	2,353	5,000		0.1	2,353	2,647
Expected Value			4,941	5,000			4,500	500
							EPD	500
							Expected Loss	5,000
							EPD Ratio	0.10

Question 3. Using the same information from the previous question, determine the beginning assets required for a 15% EPD ratio using the same algebraic calculations you would have to do on the exam (i.e. do not use a spreadsheet Solver).

Solution. Recall from the notes that this is actually quite tricky because the deficit itself depends on the asset value in a somewhat complicated way:

$$\text{Deficit} = \max(0; 5{,}000 - A)$$

which makes the algebra fairly complex.

On the exam you would have to begin by assuming that all three scenarios lead to a deficit and that the max(·) function didn't complicate things so that the deficit was simply $5{,}000 - A$. Then solve for the starting asset amount and check to make sure that all three scenarios did in fact lead to a deficit. If not, then do it again assuming that only the second and third scenarios lead to deficits and continue as before. You would have to keep doing this until you got an answer.

Let's see that in detail. Assume the starting assets are given as A. The EPD ratio is then:

$$\begin{aligned}\text{EPD Ratio} &= \frac{.1[5{,}000 - A(190.5\%)] + .8[5{,}000 - A(95.2\%)] + .1[5{,}000 - A(47.6\%)]}{5{,}000} \\ &= .15\end{aligned}$$

and then we need to solve for A. This is trivial algebra and we get $A = 4{,}250$.

However, if we use that as the starting assets, we get in Scenario 1 that ending assets equal $4{,}250(190.5\%) = 8{,}095$. But in this scenario, the deficit is zero, which means that our initial assumption that all scenarios led to a deficit was invalid and we have to try again with a different assumption.

Now assume that only Scenarios 2 and 3 lead to deficits. Here, the EPD Ratio is:

$$\begin{aligned}\text{EPD Ratio} &= \frac{.1[0] + .8[5{,}000 - A(95.2\%)] + .1[5{,}000 - A(47.6\%)]}{5{,}000} \\ &= .15\end{aligned}$$

Solving for A is then easy and gives you $A = 4{,}632$.

Checking this, we can see that the ending asset values are 8,824, 4,412 and 2,206 and the deficits are 0, 588 and 2,794. Since there is no inconsistency with our initial assumption, we are done.

The final calculations are as follows:

TABLE 7. EPD Calculation

Scenario	Beginning Asset	Return	Ending Asset	Loss	Capital	Probability	Claim Payment	Deficit
1	4,632	190.50%	8,824	5,000		0.1	5,000	0
2	4,632	95.20%	4,412	5,000		0.8	4,412	588
3	4,632	47.60%	2,206	5,000		0.1	2,206	2,794
Expected Value			4,632	5,000			4,250	750
							EPD	750
							Expected Loss	5,000
							EPD Ratio	0.15

Question 4. Suppose that an insurer has Assets at Jan 1, 2002 of \$25,000 and that at the end of the year the assets will be worth either \$20,000 or \$35,000 with probabilities .4 and .6. Also assume that their liabilities, which are uncorrelated with their assets, are due at the end of the year and will be either \$10,000, \$25,000 or \$40,000 with probabilities .25, .40 and .35. Calculate the expected policyholder deficit and the EPD ratio assuming that the risk free interest rate is zero.

Solution. First, let's set out some details so that we can follow things. We are told that the distributions for both Assets and Liabilities are discrete and uncorrelated. Therefore, it is easy to map out the joint distribution of the shortfall as follows:

TABLE 8. Shortfall Distribution

Shortfall			Probabilities		
	Assets			Assets	
Losses	20,000	35,000	Losses	20,000	35,000
10,000	0	0	10,000	0.10	0.15
25,000	5,000	0	25,000	0.16	0.24
40,000	20,000	5,000	40,000	0.14	0.21

Using these two tables, the Expected Policyholder Deficit is simply:

$$\text{EPD} = 0(.10) + 5{,}000(.16) + 20{,}000(.14) + 0(.15) + 0(.24) + 5{,}000(.21) = \$4{,}650$$

Since the expected losses equal \$26,500, the EPD Ratio is 17.5%.

Question 5. Butsic uses the Black-Scholes formula to derive the EPD Ratio in the case where the assets are lognormally distributed, the liabilities are fixed and the risk free rate is zero. To do this, he argues that the EPD for risky assets is like a put option on the Assets with a strike price equal to L, the fixed losses. Derive this formula using the Black-Scholes formula for call options and Put-Call Parity.

Solution. Following the steps shown in the Appendix, we first need to find the comparable call option on the Assets with a strike price of L. Substituting A for S and L for K in the Black-Scholes formula, and making all of the other assumptions ($r = 0, T = 1, k_A = \sigma_A$) and noting that:

$$\frac{A}{L} = \left(\frac{L}{A}\right)^{-1} = (1 - c_A)^{-1}$$

we have:

$$\text{Call} = AN(d_1) - LN(d_2)$$

where,

$$\begin{aligned} d_1 &= \frac{\ln(A/L) + (0 + \frac{1}{2}k_A^2)(1)}{k_A\sqrt{1}} \\ &= \frac{\ln(L/A)^{-1}}{k_A} + \frac{k_A}{2} \\ &= \frac{-\ln(1 - c_A)}{k_A} + \frac{k_A}{2} \\ d_2 &= d_1 - k_A \end{aligned}$$

Plugging in for d_1 and d_2,

$$\text{Call} = A\Phi\left(\frac{k_A}{2} - \frac{\ln(1 - c_A)}{k_A}\right) - L\Phi\left(-\frac{k_A}{2} - \frac{\ln(1 - c_A)}{k_A}\right)$$

Now note that $\Phi(x) = 1 - \Phi(-x)$ and use that to reverse the signs of the values in $\Phi(\cdot)$,

$$\text{Call} = A\left[1 - \Phi\left(-\frac{k_A}{2} + \frac{\ln(1 - c_A)}{k_A}\right)\right] - L\left[1 - \Phi\left(\frac{k_A}{2} + \frac{\ln(1 - c_A)}{k_A}\right)\right]$$

From Put-Call Parity we can now find the formula for the EPD:

$$\begin{aligned} \text{EPD} &= D_A \\ &= \text{Value of Put} \\ &= \text{Call} + PV(L) - A \\ &= \text{Call} + L - A \\ &= A\left[1 - \Phi\left(-\frac{k_A}{2} + \frac{\ln(1 - c_A)}{k_A}\right)\right] - L\left[1 - \Phi\left(\frac{k_A}{2} + \frac{\ln(1 - c_A)}{k_A}\right)\right] + L - A \\ &= A\left[-\Phi\left(-\frac{k_A}{2} + \frac{\ln(1 - c_A)}{k_A}\right)\right] - L\left[-\Phi\left(\frac{k_A}{2} + \frac{\ln(1 - c_A)}{k_A}\right)\right] \\ &= L\left[\Phi\left(\frac{k_A}{2} + \frac{\ln(1 - c_A)}{k_A}\right)\right] - A\left[\Phi\left(-\frac{k_A}{2} + \frac{\ln(1 - c_A)}{k_A}\right)\right] \end{aligned}$$

Then, substituting:

$$b = \frac{k_A}{2} + \frac{\ln(1 - c_A)}{k_A}$$

and noting:

$$A/L = \frac{1}{1 - c_A}$$

we can write the EPD Ratio just as in Butsic's paper:

$$\text{EPD Ratio} = \frac{\text{EPD}}{L} = \Phi(b) - \frac{\Phi(b - k_A)}{1 - c_A}$$

Question 6. Suppose you have a risky liability that is normally distributed with an expected value of \$2,000, a standard deviation of \$283 and you want to have an EPD Ratio of .001. Assume the asset values are fixed and use Butsic's formula to verify that the amount of capital required is \$584.

Solution. This is a simple application of Butsic's formula for the EPD Ratio under a normal distribution assumption. The formula given in the text is:

$$d_L = \frac{D_L}{L} = k\phi\left(-\frac{c}{k}\right) - c\Phi\left(-\frac{c}{k}\right)$$

Where,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Using the values given,

$$\begin{aligned} d_L &= .001 \\ k &= \text{coefficient of variation} = 283/2000 = .1415 \\ c &= \text{capital to expected loss ratio} = 584/2000 = .292 \\ c/k &= .292/.1415 = 2.0636 \end{aligned}$$

Then just verify that $.001 = .1415\phi(-2.0636) - .292\Phi(-2.0636) = .001$.

Question 7. Suppose you have a risky liability that is log-normally distributed with an expected value of \$2,000, a standard deviation of \$283 and you want to have an EPD Ratio of .001. Assume the asset values are fixed and use Butsic's formula to verify that the amount of capital required is \$700.

Solution. Here, the formula is:

$$d_L = \Phi(a) - (1 + c)\Phi(a - k)$$

where,

$$a = \frac{k}{2} - \frac{\ln(1 + c)}{k}$$

Also notice that,

$$a - k = -\frac{k}{2} - \frac{\ln(1 + c)}{k}$$

The value of k is the same as in the previous question but now $c = .35$. We have to verify that:

$$.001 = \Phi\left(\frac{.1415}{2} - \frac{\ln(1.35)}{.1415}\right) - (1.35)\Phi\left(-\frac{.1415}{2} - \frac{\ln(1.35)}{.1415}\right) = .001$$

Question 8. Assume that you have three risk elements with the following required capital amounts such that their EPD Ratios are all .005, Stocks: $C_A = 40$, Line 1: $C_{L1} = 200$, Line 2: $C_{L2} = 150$. If all three risk elements are fully correlated, what is the required capital to maintain the same EPD Ratio?

Solution. Here, we just take the sum of each capital requirement, so $C = 390$.

Note that I purposely used the phrase "fully correlated" which is the phrase that Butsic used when he meant that the correlation coefficients to use in his formula were 1.0, without any consideration for the adjustment due to the fact that the asset and liability items are on opposite sides of the balance sheet.

Question 9. Using the information from the previous question, assume that the three risk elements are independent and use Butsic's square root rule to determine the required capital amount.

Solution. When the risk elements are independent, the capital requirement is the square root of the sum of squares of the individual capital requirements. In this case,

$$C = \sqrt{40^2 + 200^2 + 150^2} = 253$$

Question 10. Using the information from the previous question, assume that the correlation matrix for the three risk elements is as follows and use Butsic's square root rule to determine the required capital amount.

TABLE 9. Correlation Assumptions

	A	**L1**	**L2**
A	1.0		
L1	0.2	1.0	
L2	0.6	0.5	1.0

Solution. The square root rule for non-independent risk elements calculates the total capital required as:

$$C = \sqrt{\sum C_i^2 + \sum_i \sum_{j \neq i} \rho_{ij} C_i C_j}$$

Note though that Butsic is unclear about how he handles the correlation coefficients for risk elements on opposite sides of the balance sheet. In this case, he doesn't actually use the correlation matrix as shown in the question. Instead, he reverses the signs for elements on different sides of the balance sheet. In this case, even though the correlation between losses and assets are reported as .2 and .6 for Line 1 and Line 2, the formula uses -.2 and -.6, since the assets and the liabilities are on opposite sides of the balance sheet. The correlation coefficient of .5 for the two lines of business though is used as specified.

The correlation matrix he uses is then best presented as:

TABLE 10. Adjusted Correlation Assumptions

	A	**L1**	**L2**
A	1.0		
L1	-0.2	1.0	
L2	-0.6	0.5	1.0

$$\begin{aligned} C &= \sqrt{40^2 + 200^2 + 150^2 + 2(-.2)(40)(200) + 2(-.6)(40)(150) + 2(.5)(200)(150)} \\ &= 289 \end{aligned}$$

Appendix: Derivation of Butsic's EPD Ratio for Lognormal Losses

In the notes for this reading, I showed Butsic's formulas for the EPD ratio in terms of the coefficient of variation and the capital to loss ratio in the case where the losses are assumed to be lognormal and the assets are fixed. I then described the logic for why we can think of this as a call option on the losses with an exercise price equal to the value of the assets. Now I will show you the math, which Butsic shows in the Appendix but fails to provide sufficient explanation of his steps to make it easy to understand.

Derivation Using the Black-Scholes Call Option Formula

Recall that the Black-Scholes result for a call option was:

$$\text{Call Option Value} = SN(d_1) - Ke^{-rT}N(d_2)$$

where d_1 and d_2 are defined as,

$$d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$
$$d_2 = d_1 - \sigma\sqrt{T}$$

Now, we can write the EPD using the same formula and simplifying. But there are a number of subtle tricks that Butsic uses here, so you need to be careful.

- He is assuming that $r = 0$ and that $T = 1$. In other words, he is looking at L one year from today and ignoring the effect of discounting.
- He defines L as the expected loss, but under the assumptions he is making about $r = 0$, $L_0 = E(L_T)$ so he uses L as both the current value and the expected value.
- He also assumes that k, which is the coefficient of variation, is approximately equal to σ, even though $k = \sqrt{e^{\sigma^2} - 1}$ and this is only approximate.
- He makes use of the fact that with his notation, $A = (1 + c)L$ and therefore $L/A = (1 + c)^{-1}$.
- He uses the two mathematical identities:

$$\ln(1/X) = -\ln(X) \qquad 1 - \Phi(x) = \Phi(-x)$$

Making these substitutions then, we have the following:

$$\begin{aligned} \text{EPD} &= LN(d_1) - AN(d_2) \\ &= LN(d_1) - (1 + c)LN(d_2) \end{aligned}$$

And the EPD Ratio can then be written as:

$$\text{EPD Ratio} = \frac{\text{EPD}}{L} = N(d_1) - (1+c)N(d_2)$$

where,

$$\begin{aligned} d_1 &= \frac{\ln(L/A) + (0 + \frac{1}{2}k^2)(1)}{k\sqrt{1}} \\ &= \frac{\ln[(1+c)^{-1}]}{k} + \frac{k}{2} \\ &= -\frac{\ln(1+c)}{k} + \frac{k}{2} \end{aligned}$$

$$\begin{aligned} d_2 &= d_1 - \sigma\sqrt{T} \\ &= d_1 - k \end{aligned}$$

Plugging in for d_1 and d_2,

$$\begin{aligned} \text{EPD Ratio} &= \Phi\left(-\frac{\ln(1+c)}{k} + \frac{k}{2}\right) - (1+c)\Phi\left(-\frac{\ln(1+c)}{k} + \frac{k}{2} - k\right) \\ &= \Phi(a) - (1+c)\Phi(a-k) \end{aligned}$$

where,

$$a = \frac{k}{2} - \frac{\ln(1+c)}{k}$$

Cummins: Allocation of Capital in the Insurance Industry

Role of Capital

Capital is held by insurers to assure policyholders that claims will be paid, even if larger than expected. Without such protection, insureds would be unable to hedge the credit exposure and could be left with no real insurance protection.

Capital Allocation

Why are insurers concerned with the allocation of their total capital to specific lines of business or products? It's a rather odd concept, given that all of the firm's capital stands behind all of the firm's liabilities — the firm cannot go bankrupt line by line.

Nonetheless, Cummins sees three reasons to allocate capital:

- Pricing, underwriting and other decision making could possibly be enhanced by thinking of capital as being allocated, even though it is not literally allocated.
- Allocation of capital could help to tie together certain financial decisions and regulatory risk-based capital rules.
- Concepts like risk-adjusted return on capital (RAROC, discussed in the Goldfarb paper) and economic value added (EVA) make use of capital allocation for performance measurement.

Value Maximization

Cummins adopts the standard perspective that the goal for management should always be to maximize the shareholder value of the firm. He argues that in order to do this, economic (present value) profits must be weighed against the capital requirement (a risk measure) for the businesses.

There are two approaches discussed, RAROC and EVA.

RAROC — Risk-Adjusted Return on Capital

One way to do this is by first allocating the total capital of the firm (or some portion of it) to various businesses, presumably in such a way as to capture the relative riskiness of each business. In particular, we should be sure to capture the *contribution* that each business makes to the total riskiness of the firm, reflecting positive or negative correlations as appropriate. Then, the after-tax present value net income can be compared to the allocated capital for that business and a ratio, termed the *risk-adjusted return on capital*, or RAROC, can be calculated.

Cummins then argues that to maximize shareholder value, the RAROC for each business should be compared to a target measure. A RAROC above this target implies that shareholder value is being increased; a RAROC below this target implies that shareholder value is being reduced.

EVA — Economic Value Added

A closely related approach known as economic value added does not calculate a ratio of income to capital as in RAROC. Instead, the income is reduced by the product of the allocated capital and the target rate (the cost of capital as that term is often used). If the resulting EVA is positive then this implies value is being increased; if the EVA is negative value is being reduced.

Important Note Regarding Cost of Capital

Notice that both the RAROC and the EVA approaches make use of the cost of capital to reflect the fair return on risk-based capital that must be earned in order to create value. Where does this come from? Cummins simply states that it is derived using "an appropriate asset pricing model". I argue that it is far more complex than this, but for the present purposes we will leave it at that. Just note that he goes on to state that it could be estimated using the returns shareholders expect from other firms in the same business, using "pure plays" that operate in only the business of interest. Alternatively, conglomerate firms in many businesses could be used to infer the returns for each of their component businesses, using a method known as the full information beta approach.

Capital Allocation Techniques

The rest of the paper focuses on the methods used to allocate capital to different businesses. The key point is that the firm will operate with a certain amount of capital depending on the overall solvency risk of the firm. An allocation of this capital is then done to ensure that the capital is attributed to the businesses that create the risk in the first place.

The methods discussed are listed here and explained in more detail below:

- Regulatory (NAIC) Risk-Based Capital
- CAPM
- Value at Risk
- Insolvency Put Option (Expected Policyholder Deficit)
- Merton-Perold
- Myers-Read

Regulatory (NAIC) Risk-Based Capital

The NAIC developed a model to determine firmwide capital requirements based in part on an assessment of the risks faced from a variety of risk sources. Briefly stated, the NAIC approach contains charges applied to various balance sheet and income statement items to measure the risk in 6 categories:

- R_0 — Risk associated with equity holdings of subsidiaries
- R_1 — Risk of credit-related losses in bond portfolio and changes in value in equity portfolio and other investments
- R_2 — Risk of loss reserves developing adversely
- R_3 — Risk of losses on new business written during the coming calendar year
- R_4 — Credit risk associated with agents' balances and reinsurance receivables
- R_5 — Off balance sheet risks

After these charges are determined, they are combined using a square-root rule that essentially assumes zero correlation among the R_1 to R_5 risks and perfect correlation of this with the R_0 risks. This is, of course, a crude approximation.

CAPM

Cummins lists CAPM as a capital allocation technique, though it isn't really used to allocate capital and so the discussion is somewhat out of place.

Instead, Cummins argues that CAPM is used to determine each line's contribution to the total firmwide required profits by decomposing the firmwide CAPM beta into a component reflecting the asset risk and components for each line of business. Then CAPM is used to measure the required return for each business based on its implied beta.

There are a host of issues associated with the use of CAPM in this fashion. To begin, CAPM is a model of systematic risk from the perspective of a diversified investor. Capital requirements for insurers are driven by policyholders who are not diversified in the same fashion and therefore may be concerned with line-specific tail risks in addition to systematic risk. Secondly, estimating CAPM-type betas by line of business is very difficult to do given data limitations. Thirdly, CAPM may not even be a good model for equity prices, let alone businesses within the firm. Recall that models such as APT or the Fama French Three-Factor model suggest that there may be other priced risk factors besides CAPM's market risk.

With all of those caveats in mind, let's focus on the formula presented in the reading. Start by assuming a company writes one line of business and decompose the firm's return on equity (surplus) into investment income on its assets and underwriting income.

$$r_E = r_A(1 + ks) + r_U s$$

where r_A is the investment return, s is the ratio of premium to equity and k is the proportion of premium that represents policyholder funds that can be invested and r_U is the underwriting profit relative to premium.

That's a rate of return, which we could also write in terms of CAPM betas:

$$\beta_E = \beta_A(1 + ks) + \beta_U s$$

Now we can use CAPM to plug in for r_A in the first equation:

$$r_E = [r_f + \beta_A(r_m - r_f)](1 + ks) + r_U s$$

Similarly, we can plug the formula for β_E into CAPM to get another expression for r_E:

$$r_E = r_f + [\beta_A(1 + ks) + \beta_U s](r_m - r_f)$$

Now, just set these two equations for r_E equal to each other and solve for r_U:

$$r_U s = -r_f ks + \beta_U s(r_m - r_f)$$
$$r_U = -kr_f + \beta_U(r_m - r_f)$$

We could easily extend the above logic to handle a firm with multiple lines of business.

Note that in the Cummins reading, the above is presented differently. In the reading, k is defined as the ratio of reserves to equity (surplus) rather than as the ratio of reserves to premium. I believe this is just an error. It was, however, tested in 2012 and the CAS expected you to use the formula as it was presented in Cummins.

Value at Risk

Cummins reviews the concept of VaR and suggests that capital could be allocated by determining the *stand-alone* VaR for each business at the same selected percentile. The stand-alone VaR identifies the amount of assets, in addition to the assets sufficient to pay the expected claims, that would produce a specific probability of "default". Here, default is defined as having insurance claims in excess of these total assets.

Leaving aside the observation made earlier that a line of business cannot default on its own, there are three issues associated with VaR as a mechanism for the allocation of capital.

1. Firms may not have enough capital in total to equalize the probability across all lines of business for any given excedence probability.

2. It is difficult to reflect the diversification effect in this approach — it uses the stand-alone VaR amounts.

3. As discussed in Butsic's EPD paper, the VaR approach ignores the amount by which the losses exceed the available capital.

Insolvency Put Option (EPD)

Cummins reviews Butsic's EPD methodology and then argues that capital could be allocated by equating the EPD Ratios (the EPD as a percent of the expected value of the liabilities), choosing an arbitrary target ratio such as 5%. An example of his calculations is provided in the questions below.

However, be careful to note that the Cummins notation differs slightly from the Butsic notation. Recall from Butsic that the interpretation of the EPD as a put is based on the traditional notion of fixed liabilities and risky assets. In that case, the EPD is just a put option on the assets with a strike price equal to the value of the liabilities. When the reverse is true (fixed assets and risky liabilities) the EPD is more accurately a call option on the liabilities with a strike price equal to the assets. And in general, when both assets and liabilities are risky, the option is more complicated — it is technically an option to *exchange* the assets for the liabilities.

In the Cummins text, he is specifically dealing with the case of risky assets and risky liabilities. Therefore, it can be confusing whether to model the EPD as a put on the assets or a call on the liabilities. Since Butsic's formulas for the EPD Ratio differed depending on which case he was working with, it isn't clear which of these should be used.

To generate the results in Figure 2 of the paper and in the numerical examples, Cummins actually made it easier by working with the asset to liability *ratio* rather than either the assets or the liabilities. Then he was able to calculate the *EPD Ratio* using the standard Black-Scholes put formula.

So in Cummins, the EPD Ratio is a standard put option on the *ratio* of the assets to the liabilities with a strike price equal to 1.0. If the ratio of assets to liabilities is less than 1.0, then the firm is insolvent and the deficit (as a percent of the liabilities) is the amount by which the ratio is below 1.0. The volatility parameter in the Cummins examples is therefore the volatility of the asset-to-liability ratio. It reflects not just the separate asset and liability standard deviations but also their correlations.

In the numerical problems shown later, I will use a simplifying case of fixed assets and random liabilities in order to facilitate the use of Butsic's formulas. This is an easy way to review those formulas and also makes it easier to emphasize the key concepts from Cummins.

Cummins notes that using the EPD Ratio approach is an improvement over the VaR approach, but it still fails to take diversification into account. The next two methods discussed in the Cummins paper address this shortcoming.

Note that the values shown in this section of the Cummins paper and in Figure 2 appear to be incorrect due to rounding.

Marginal Capital Allocation — Merton & Perold

The first of two marginal capital allocations methods discussed is the Merton & Perold (M-P) method.

The M-P method is based on their concept of *risk capital*, which they define as the smallest amount that can be invested in risk free assets to ensure that they do not default on their obligations (note that I defined it differently than in the text). So in some sense, risk capital by their definition is the cost of insurance against the net asset and liability value fluctuations. The catch is that sometimes some of this capital is provided by the liability holders themselves and does not always have to be supplied in the form of cash by the firm's equity holders.

In the Cummins discussion, the concept of risk capital as the basis for capital allocation is equivalent to an EPD Ratio standard. The only distinction between the M-P approach and the EPD approach discussed earlier is the fact that the M-P approach looks at the marginal impact of each line of business by including or excluding the line in its entirety.

They begin by calculating the required capital for the total firm, taking into account the diversification of risk from the various businesses. Then, for each line of business they calculate the required capital assuming that the firm consists of all businesses except the particular line and similarly calculate the required capital. The difference between these two capital amounts is the amount of marginal capital required when this particular line of business is added to all of the other businesses.

One subtle point to note about the example in Cummins is that the σ parameter reflects the coefficient of variation for the asset to liability ratio. In order to get the firmwide σ for a multi-line firm, he converts the variability measure into standard deviation in dollars by multiplying the σ by the expected losses, derives the overall standard deviation in dollars using the correlation assumptions and then converts back to a coefficient of variation by dividing by the expected aggregate losses. See the questions below for a simple example similar to what is shown in the text.

Marginal Capital Allocation — Myers-Read

The second marginal approach discussed is the Myers-Read approach. This is very similar to the Merton-Perold approach except that the Merton-Perold approach adds or subtracts entire businesses to get what Cummins refers to as a macro marginal allocation. In the Myers-Read method, they examine the marginal capital for a particular line of business by determining the effect of a small increase in the size of the line (based on expected loss amount).

The math to accomplish this is fairly messy. It involves writing the formula for the insolvency put (the EPD) as a function of the loss amounts for each line and calculating partial derivatives with respect to each line. The capital needed for each line is then determined such that each line has the same marginal impact on the firm's overall insolvency put value (as a percentage of the line's expected losses).

Notice a subtle aspect of Cummins' example. The Myers-Read method is designed to allocate a set amount of capital and so it does not necessarily tell you how much that total capital amount should be. In the Cummins example, he assumes the same 5% EPD Ratio target for the firm overall that he used in his Merton-Perold example is appropriate. His Myers-Read allocation simply allocates this total amount.

A specific formula is given in the text for the resulting surplus to liability ratio for each line assuming that each line has the same marginal default value as the firm overall.

$$s_i = s - \left(\frac{\partial p}{\partial s}\right)^{-1}\left(\frac{\partial p}{\partial \sigma}\right)[(\sigma_{iL} - \sigma_L^2) - (\sigma_{iV} - \sigma_{LV})]\frac{1}{\sigma}$$

In that formula, the σ terms with the V subscripts reflect the covariance of the line i losses with the assets and the covariance of the total losses L with the assets, respectively. The formula shown is the general formula in the case of risky assets and risky liabilities, so in the case of the example used in the reading it can be simplified a bit. We will do this in the Practice Questions.

In the end, the Myers-Read approach is more appealing than the Merton-Perold approach because the sum of the marginal capital allocations is equal to the total capital requirement for the firm. There is no need to worry about the unallocated capital. This *summing up* quality is appealing in many contexts.

Note for Interested Students — Steve Mildenhall has written an excellent paper discussing a significant shortcoming of the Myers-Read method. It is available on the CAS website.

Economic Cost of Capital vs. Cost of Capital

Earlier in the text Cummins used the term *cost of capital* to refer to the return per unit of capital that the investors require given the systematic risk of the business. He used CAPM (or some other asset pricing model) to determine this amount. In Section 3.7, he discusses the *economic cost of capital*, a different concept altogether.

The economic cost of capital reflects the actual costs associated with an insurance company holding capital. These are the actual costs that might be incurred by the shareholders when the firm holds capital and simply invests it in marketable securities. The costs come from three main sources:

- Agency Costs — These are the costs incurred when management has too much cash burning a hole in its pocket and wastes it on things that do not add value for shareholders, such as excessive compensation, perks such as corporate jets and fancy offices, ego-driven growth and acquisition strategies, etc. Cummins also includes sloppy pricing and underwriting in this category as well, which becomes easier to do when the company is flush with capital.

- Double Taxation — In some jurisdictions, investment income earned by an insurance company is taxed when it is earned and then taxed again when it is paid to shareholders in the form of dividends. This creates a double taxation cost to shareholders who might otherwise be able to invest the capital in the same marketable securities and only be taxed once.

- Regulatory Costs — Regulatory restrictions on investments may result in inefficient and sub-optimal investment of the capital.

Each of these costs can be considered the frictional costs of holding capital. As a result, a business whose riskiness results in the need to hold more capital will need to earn enough profits to not just cover the profits demanded by shareholders (the CAPM-based cost of capital reflecting systematic risk) but also the frictional costs incurred. Of course, quantifying these costs is quite tricky, but at least in theory they need to be reflected.

Conclusions of the Cummins Paper

The following are the key conclusions of the paper:

- EPD is better than VaR, even though both might be useful to calculate. He also thinks it is better to estimate EPD and VaR at different thresholds rather than at just a single point.

- The option models are better than EPD or VaR because they allow recognition of diversification benefits and he prefers the Myers-Read to the Merton-Perold method.

- The economic cost of capital (the frictional cost or the spread cost) is what should be allocated to the lines of business.

- Capital allocation must reflect both the asset and liability risks and in particular the covariance between them.

- The duration and maturity of the liabilities should be reflected in the capital allocation.

- The decision making system should dictate the data needs, not the other way around.

- Capital allocation will lead to better pricing, underwriting and strategy decisions and will lead to shareholder value creation for the winning firms.

Practice Questions

Question 1. Cummins argues that a major motivation for capital allocation by insurers is so that they can properly assess whether certain businesses are earning enough profits to compensate the firm's shareholders for the risk to their capital. What are the two performance measurement approaches briefly outlined by Cummins?

Solution. The two approaches he describes are risk-adjusted return on capital (RAROC) and economic value added (EVA).

Question 2. List the 6 methods Cummins describes for allocating capital.

Solution. The methods discussed are:

- Regulatory (NAIC) Risk-Based Capital
- CAPM
- Value at Risk
- Insolvency Put Option (Expected Policyholder Deficit)
- Merton-Perold
- Myers-Read

Question 3. What are the reasons given for why he does not advise using the NAIC's RBC method to allocate capital? What two reasons does he give for why the NAIC RBC approach is still important to understand?

Solution. The weaknesses he identified in his introductory remarks included:

- the charges are inaccurate,
- they are based on book values rather than market values,
- they ignore some important risk sources such as interest rate risk and certain derivatives exposures,
- they are based on industry data for the typical insurer.

Later, he also noted the following additional weaknesses:

- there was little theoretical foundation for the NAIC model,
- the reserve and underwriting charges reflect worst case scenarios rather than variances,
- correlations among the firm's businesses was not adequately addressed.

Nonetheless, the framework the NAIC used does reflect most of the risk sources and the regulatory capital requirements do serve as legitimate constraints regardless of the results of other approaches.

Question 4. Suppose a line of business had losses with an expected value of $1,000 and a standard deviation of $300, implying a lognormal distribution with parameters $\mu = 6.865$ and $\sigma = .2936$. Another line of business also has a lognormal distribution with a mean of $1,000 and a standard deviation of $500, implying parameters $\mu = 6.796$ and $\sigma = .4724$. All assets are fixed. Further assume that you wanted to use an EPD ratio of 5% to allocate capital to these two lines following the approach Cummins used. What would be the capital allocated to each line?

Solution. Recall from Butsic's EPD paper that the formula for the capital ratio, c, is rather complicated and cannot be easily solved for. However, it is possible to use his formula for the EPD ratio as a function of the lognormal sigma and the capital ratio to build a table that can then be used to look up the 5% EPD ratio level and linearly interpolate to find the appropriate c value.

For the sake of review, recall that Butsic derived the following formula for the EPD ratio in the case of fixed assets and risky liabilities by assuming the following:

$$L = E(L), r = 0, c = \text{Capital}/L, T = 1$$

$$\text{EPD Ratio} = \Phi(a) - (1 + c)\Phi(a - k)$$

where,

$$a = \frac{k}{2} - \frac{\ln(1 + c)}{k}$$

Using this formula, but with the actual lognormal σ parameter rather than Butsic's approximation using the coefficient of variation (k), the following tables can be constructed for the EPD ratio at different values for c and with two different values for σ. Note that this table is similar to the graph shown in Cummins' Figure 2, except that I show the capital ratio rather than the asset to liability ratio, which is just $1 + c$.

TABLE 1. Approximate Capital Ratio and EPD Ratio

Sigma = 29.36%		Sigma = 47.24%	
c	EPD Ratio	*c*	EPD Ratio
50.00%	1.36%	80.00%	3.18%
47.50%	1.53%	77.50%	3.36%
45.00%	1.72%	75.00%	3.55%
42.50%	1.93%	72.50%	3.75%
40.00%	2.16%	70.00%	3.96%
37.50%	2.42%	67.50%	4.19%
35.00%	2.70%	65.00%	4.42%
32.50%	3.02%	62.50%	4.67%
30.00%	3.38%	60.00%	4.94%
27.50%	3.77%	57.50%	5.22%
25.00%	4.20%	55.00%	5.52%
22.50%	4.68%	52.50%	5.83%
20.00%	5.21%	50.00%	6.16%
17.50%	5.79%	47.50%	6.51%
15.00%	6.43%	45.00%	6.89%

Based on these tables, the capital ratio for the first line can be linearly interpolated as 21.0%, suggesting a capital allocation of:

$$C_1 = 21.0\%(\$1{,}000) = \$210$$

This is in addition to the capital required to support the expected loss of $1,000, which means that the total assets required are $1,210.

For the second line, the capital ratio is 59.4%, implying a capital allocation of $594 and total assets of $1,594.

Note that we can verify these calculations by calculating the EPD Ratios directly given the result that we need assets equal to either $1,210 or $1,594. Recall that the EPD when the liabilities are risky and the assets are fixed is a call option on the liabilities with the following parameters and call option values:

TABLE 2. Verification of EPD Ratios

	Line 1	Line 2
Spot (Liability Value)	1,000	1,000
Strike (Asset Value)	1,210	1,594
Volatility	29.36%	47.24%
Risk Free Rate	0%	0%
Time to Expiration	1	1
EPD = $LN(d_1) - Ae^{-rT}N(d_2)$	50	50
EPD Ratio	5%	5%

Question 5. Using the information from the previous question, assume you wanted to determine the allocated capital for each line of business using the Merton-Perold method and the same 5% EPD Ratio as the basis for the capital determination. If the two lines of business have a 70% correlation coefficient, what would you use as the volatility parameter using the same method as in Cummins?

Note: This question is not really important from the perspective of the learning objectives. You should feel free to skip to the next question, which uses the result from this question. The details are here only for those of you struggling to recreate the values in Cummins' examples.

Solution. The volatility parameter that Cummins uses is technically the standard deviation of the natural log of the asset to liability ratio, A/L. In the case where the assets are fixed and the liabilities are risky, this turns out to be the same for a single line of business as the standard deviation of the log of the liability, as shown below:

$$\text{Var}\left[\ln\left(\frac{A}{L}\right)\right] = \text{Var}[\ln(A) - \ln(L)] = \sigma_L^2$$

because $\sigma_A = 0$ in this case.

In the previous question, you were given the σ parameter of the lognormal liability distribution, which is also the volatility of the asset-to-liability ratio.

However, in this question, you need to determine the volatility of the asset-to-liability ratio when there is more than one line of business. What Cummins appears to do is to calculate the total firmwide volatility parameter as follows:

$$
\begin{aligned}
w_1 &= \frac{L_1}{L_1 + L_2} = \frac{1{,}000}{1{,}000 + 1{,}000} = \frac{1}{2} \\
w_2 &= 1 - w_1 = \frac{1}{2} \\
\sigma_T^2 &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho \\
&= (.5)^2(.2936)^2 + (.5)^2(.4724)^2 + 2(.5)(.5)(.2936)(.4724)(.7) \\
&= .1259 \\
\sigma_T &= .3548
\end{aligned}
$$

Question 6. Using the information from Questions 4 & 5, determine the allocated capital for each line of business using the Merton-Perold method and a 5% EPD Ratio as the basis for the capital determination. Recall from Question 4 that the stand-alone capital requirements for Line 1 and Line 2 were \$210 and \$594 and the firmwide volatility from Question 5 was $\sigma_T = .3548$.

Solution. This is a simple two-line version of the example given in the text and should clarify the approach with fewer calculations.

Creating a table of the firmwide EPD Ratio at different levels of the capital ratio, we can again solve for the required capital for the combined firm.

Sigma = 35.48%	
c	EPD Ratio
50.00%	2.71%
47.50%	2.95%
45.00%	3.22%
42.50%	3.50%
40.00%	3.82%
37.50%	4.16%
35.00%	4.52%
32.50%	4.92%
30.00%	5.35%
27.50%	5.82%
25.00%	6.33%
22.50%	6.87%
20.00%	7.46%
17.50%	8.10%
15.00%	8.78%

Using this table and interpolating to find the c associated with a 5% EPD Ratio, we get c = 32.05%. This translates into a total capital requirement for the combined firm of:

$$C_T = 32.05\%(\$2{,}000) = \$641$$

To find the Merton-Perold marginal capital requirements, notice that if the firm had only consisted of line 2 then the capital requirement would have been \$594, as we found in Question 4. When both lines exist, the capital is \$641 according to the calculations here. This suggests that the *marginal* capital from adding line 1 is \$641 − \$594 = \$46.

Similarly, had we started with line 1 the capital required would have been \$210. The marginal capital required when we add line 2 is therefore \$641 − \$210 = \$431.

Using these marginal capital requirements, we would allocate \$46 of capital to line 1 and \$431 of capital to line 2.

The sum of the marginal capital amounts is only \$46 + \$431 = \$477, which is lower than the \$641 capital required for the combined firm. The rest of the firm's total capital would not be allocated to any particular line and would instead be classified as corporate capital.

Question 7. The Merton-Perold and Myers-Read method are both described as marginal capital allocation methods and both rely on the insolvency put value (EPD). What is the major difference between the two methods? Which does Cummins think is the more appealing of the two?

Solution. The Merton-Perold approach measures the impact on the put value of completely adding or removing an entire business whereas the Myers-Read method measures the impact of a small change in the size of the business. The effect of this difference is that the marginal capital amounts under Merton-Perold do not add up to the total capital requirement under the same standard (i.e. the same target EPD ratio). The Myers-Read method does lead to additive capital allocations. Since most decisions are more likely to involve adding or removing small amounts of a business (writing one more policy, non-renewing one account) Cummins seems to prefer the Myers-Read method.

Question 8. List the three sources of frictional costs associated with holding capital in an insurance company.

Solution. The three given in the text are:

- Agency costs associated with management not acting in the interest of shareholders,
- double taxation of investment income on capital,
- regulatory costs associated with investment restrictions and other regulatory constraints.

Question 9. List the major conclusions of the Cummins paper.

Solution. The following are the key conclusions of the paper:

- EPD is better than VaR, even though both might be useful to calculate. He also thinks it is better to estimate EPD and VaR at different thresholds rather than at just a single point.
- The option models are better than EPD or VaR because they allow recognition of diversification benefits and he prefers the Myers-Read to the Merton-Perold method.
- The economic cost of capital (the frictional cost or the spread cost) is what should be allocated to the lines of business.
- Capital allocation must reflect both the asset and liability risks and in particular the covariance between them.
- The duration and maturity of the liabilities should be reflected in the capital allocation.
- The decision making system should dictate the data needs, not the other way around.
- Capital allocation will lead to better pricing, underwriting and strategy decisions and will lead to shareholder value creation for the winning firms.

Question 10. Use the Myers-Read method to allocate the capital from Question 5, assuming that the firmwide capital is set to achieve the same EPD Ratio target as in that question.

Note — This is a difficult question that requires you to use the Myers-Read formula. It ought to have been considered beyond the scope of the learning objectives, but it was asked on the 2010 exam.

Solution. The first thing to note is that this question involves risky liabilities and fixed assets, so the default option is valued as a simple call option, which makes it easy to determine the partial derivatives needed for the Myers-Read method.

Let's write the formula Cummins uses for the resulting surplus to liability ratios, s_i:

$$s_i = s - \left(\frac{\partial p}{\partial s}\right)^{-1} \left(\frac{\partial p}{\partial \sigma}\right) [(\sigma_{iL} - \sigma_L^2) - (\sigma_{iV} - \sigma_{LV})] \frac{1}{\sigma}$$

In that formula, the σ terms with the V subscripts reflect the covariance of the line i losses with the assets and the covariance of the total losses with the assets, respectively. Since the assets were assumed to have no volatility, both of those terms are zero in our case.

In addition, the σ parameter with no subscripts reflects the overall volatility of the assets and liabilities. Again, with no asset volatility, that is just the volatility of the liabilities, σ_L.

Finally, the term σ_{iL} reflects the covariance of line i with the total losses for all lines. Using the expression for the total variance, we can write the relationship between this covariance term and the total variance as follows:

$$\begin{aligned}
\sigma_L^2 &= \sum_i w_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j \rho \\
&= \sum_i \sum_j w_i w_j \sigma_i \sigma_j \rho \\
&= \sum_i w_i \left[\sum_j w_j \sigma_i \sigma_j \rho \right] \\
&= \sum_i w_i \left[\sigma_{iL} \right]
\end{aligned}$$

For the partial derivative terms, we begin by writing the formula for expected default value as a call option on the liabilities (L) with a strike price equal to the current value of the assets (A). We'll keep the notation simple by assuming $r = 0$ and $T = 1$.

$$\text{Insolvency Put} = LN(d_1) - AN(d_2)$$

where,

$$\begin{aligned}
d_1 &= \frac{\ln(L/A) + \frac{1}{2}\sigma^2}{\sigma} = -\frac{\ln(1+s)}{\sigma} + \frac{\sigma}{2} \\
d_2 &= d_1 - \sigma
\end{aligned}$$

Notice that I left out the terms for the interest rate, since it is assumed to be zero, and the time to maturity, since it is assumed to be 1.0. I also substituted in $1 + s = A/L$.

Now, if we define the default value as a percentage of the liabilities and divide the formula above by L, we can rewrite the default ratio as:

$$p = N(d_1) - (1+s)N(d_2)$$

From this it is relatively easy to calculate the two partial derivatives we need for our formula for the surplus to liability ratio. We take the derivative of p with respect to s but pay attention to the fact that there is an s term in the d_1 and d_2 terms that makes the derivative a bit of a mess. (Interested students can download a file containing these derivations from GoldfarbSeminars.com).

However, the messy formulas simplify so that:

$$\frac{\partial p}{\partial s} = -N(d_2)$$

$$\frac{\partial p}{\partial \sigma} = N'(d_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_1^2}$$

Calculating all of these values is then relatively easy and the result is shown below:

$$
\begin{aligned}
\sigma_{1L} &= .5(.2936)(.2936)1 + .5(.2936)(.4724)(.7) = .0916\\
\sigma_{2L} &= .5(.2936)(.4724)(.7) + .5(.4724)(.4724)(1) = .1601\\
\sigma = \sigma_L &= \sqrt{.5(.0916) + .5(.1601)} = \sqrt{.1259} = .35478\\
s &= 641/2000 = .3205\\
d_1 &= -\frac{\ln(1+s)}{\sigma} + .5\sigma = -.60624\\
d_2 &= d_1 - \sigma = -.9610\\
p &= N(d_1) - (1+s)N(d_2) = 5\%\\
\frac{\partial p}{\partial s} &= -N(d_2) = -.16827\\
\frac{\partial p}{\partial \sigma} &= N'(d_1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(.60624)^2} = .33197
\end{aligned}
$$

$$
\begin{aligned}
s_1 &= .3205 - (-.16827)^{-1}(.33197)(.0916 - .1259)\frac{1}{.35478} = 13.0\%\\
s_2 &= .3205 - (-.16827)^{-1}(.33197)(.1601 - .1259)\frac{1}{.35478} = 51.1\%
\end{aligned}
$$

And then in terms of dollars, the allocation to lines 1 and 2 are:

$$C_1 = s_1 L = 13.0\%(1{,}000) = \$130$$

$$C_2 = s_2 L = 51.1\%(1{,}000) = \$511$$

Notice that the total allocation is \$641, proving that the marginal allocations add up to the total capital of the firm. Notice also that $p = 5\%$, which is the target EPD ratio for the firm overall.

Question 11. In Questions 4 and 5, I used a special case with fixed assets and risky liabilities and made use of Butsic's simplified formulas for the EPD Ratio as a function of the capital ratio, c. In Cummins, he suggests that when the assets and liabilities are both risky the EPD Ratio can be found simply by using the standard Black-Scholes put formula. What assumption does he make in order to do this?

Solution. Cummins assumes that the asset-to-liability ratio is lognormally distributed with a volatility parameter of σ. This allows him to write the EPD ratio as a standard put on the asset-to-liability ratio $(1 + c)$ with a fixed strike price equal to 1.0.

Although he writes the formula for the "EPD" as a standard Black-Scholes put using the notation:

$$P(A, L, r, t, \sigma)$$

as he shows in Footnote 4 he is actually calculating the *EPD Ratio* using the formula:

$$\text{EPD Ratio} = P(A/L, 1.0, r, t, \sigma)$$

In that formula, A/L is the asset-to-liability ratio and 1.0 is the strike price. The σ parameter is the standard deviation of the natural log of the asset-to-liability ratio.

Since the formula has risky underlying "assets" (the A/L ratio actually) and a fixed strike price, the model can now use the standard Black-Scholes put formula or Butsic's formulas, whichever is easier.

Question 12. Consider a firm with \$4,427 of total assets and \$3,000 in expected liabilities. The asset-to-liability ratio in this case is 4,427/3,000 = 1.476. If the volatility of the asset-to-liability ratio is 42.59%, what is the EPD Ratio using the standard Black-Scholes formula for a put? Assume that the time to maturity is 1.0 and the risk free rate is zero.

Solution. Using Put-Call Parity and the standard Black-Scholes formula for a call option on the asset-to-liability ratio with a strike price of 1.0, the EPD Ratio is as follows:

$$\begin{aligned}
\text{Put} &= \text{Call} + \text{PV(Strike)} - \text{Stock} \\
&= \text{Call} + \text{PV(1.0)} - \text{Asset-to-Liability Ratio} \\
\text{Call} &= (\text{Asset-to-Liability Ratio})N(d_1) - 1.0N(d_2) \\
d_1 &= \frac{\ln(\text{Asset-to-Liability Ratio}/1.0) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \\
d_2 &= d_1 - \sigma\sqrt{T}
\end{aligned}$$

Plugging in, we have:

$$\begin{aligned}
d_1 &= \ln(1.476)/.4259 + .4259/2 = 1.126 \\
d_2 &= .7004
\end{aligned}$$

This leads to:

$$\text{Call} = (1.476)N(1.126) - N(.7004) = .5256$$

Finally, using put-call parity the EPD Ratio is:

$$\text{EPD Ratio} = .5256 + 1.0 - 1.476 = 0.05$$

Note that this numerical example is identical to the example shown in the text, where the EPD Ratio was set to 5% and the required capital was found to equal \$1,427. I merely made you solve it the other way around so that I could clarify Footnote 4 in Cummins. In Cummins the text suggests that this is the formula for the EPD, when in fact it is actually the formula for the EPD Ratio.

Notice that I also told you to use the asset-liability ratio defined as the ratio of the current assets to the expected liability (paid at the end of the year) and to assume that the risk free rate is zero. This was done merely to remain consistent with the Cummins paper, which also assumed the risk free rate was zero. See my solution to 2007 Exam Question 36 for a more complete discussion of the importance for consistency between the values used for the liability and the risk free rate.

Question 13. Redo the previous calculation using Butsic's formula for the EPD Ratio when the assets are risky and the liabilities are fixed.

Solution. Recall that in Butsic, he defined the variable c_A such that $A/L = (1 - c_A)^{-1}$. He then provided the following formula for the EPD Ratio when assets were risky and the liabilities were fixed:

$$\text{EPD Ratio} = \Phi(b) - \frac{\Phi(b - k_A)}{1 - c_A}$$

where,

$$b = \frac{k_A}{2} + \frac{\ln(1 - c_A)}{k_A}$$

Note also that $A/L = 1/(1 - c_A)$ and k_A is the coefficient of variation for the asset-to-liability ratio. Butsic used the coefficient of variation as an approximation of the volatility (standard deviation) of the asset-to-liability ratio. Here we have the actual volatility parameter and so we will use that instead of Butsic's approximation.

Here, $c_A = .322$ and $b = .4259/2 + \ln(1 - .322)/.4259 = -.7004$. This gives a final EPD Ratio of:

$$\begin{aligned}
\text{EPD Ratio} &= \Phi(b) - \frac{\Phi(b - k_A)}{1 - c_A} \\
&= N(-.7004) - \frac{N(-.7004 - .4259)}{1 - .322} \\
&= 0.05
\end{aligned}$$

Notice that this is the same as the previous formula that used the Black-Scholes formula directly. The reason this works is that everything is expressed in terms of the liabilities and so on that basis, the strike price is indeed *fixed* at 1.0.

Goldfarb: Risk-Adjusted Performance Measures for P&C Insurers

This paper includes a lot of commentary and some details that are not addressed in the learning objectives. Therefore, I will extract only the portions of the paper that seem the most relevant for the exam in these notes and encourage you to read the full paper for further details.

RAROC Definition

Risk-adjusted performance measures are intended to improve upon the metrics used to make capital planning, risk management and corporate strategy decisions by explicitly reflecting the risks inherent in different businesses.

In a simple one-period case in which a business requires an investment of a specific amount of capital and earns (or is expected to earn) a given dollar amount of income (profit) during the period, the return on capital is simply calculated as:

$$\text{Return on Capital} = \frac{\text{Income}}{\text{Capital}}$$

This is, of course, a very general form of a return calculation and in practice there are a wide variety of approaches that can be used to determine the amounts used for both the numerator and denominator.

Many banks and insurance companies have adopted *risk-adjusted* return on capital measures in which either the return is risk-adjusted, the capital is risk-adjusted, or in some cases both are risk-adjusted. Often all three instances are generically referred to as RAROC (Risk Adjusted Return On Capital), a convention that will be used here for convenience. But for clarity, throughout this discussion the emphasis will be on a risk-adjusted return measure where the income is not risk-adjusted and the capital is risk-adjusted:

$$\text{Risk-Adjusted Return on Capital} = \frac{\text{Income}}{\text{Risk-Adjusted Capital}}$$

Alternative Income Measures for RAROC

A wide variety of income measures exist, all of which are intended to reflect the profit, in dollars, during a specific measurement period. Four relevant choices include:

- GAAP Net Income
- Statutory Net Income
- IASB Fair Value Basis Net Income
- Economic Profit

GAAP Net Income

This measures the income earned according to GAAP accounting conventions. Use of this measure is convenient when RAROC is intended to be used to guide management decision-making, since the measurement basis is already in use within the firm.

Statutory Net Income

In countries where separate statutory (regulatory) accounting frameworks are used, the income component may also be measured using these statutory accounting conventions.

IASB Fair Value Basis Net Income

Although not yet formally adopted, efforts are underway by the International Accounting Standards Board (IASB) to develop *fair value* accounting standards. These standards are intended to remove many existing biases in various accounting conventions used throughout the world. For insurance companies, this measure of profit differs from GAAP net income primarily due to the discounting of loss reserves to reflect their present value and the inclusion of a risk margin on loss reserves to approximate a risk charge that would typically be included in an arms-length transaction designed to transfer the risk to a third party.

Economic Profit

A more general method for measuring profit that further eliminates many accounting biases is often referred to as *economic profit*. Unfortunately, this term is often used to refer to many different types of adjustments to the GAAP income measures. Generally it refers to the total change in the economic value of the assets and liabilities of the firm, where asset values reflect their market value and the liabilities are discounted to reflect their present value. Whether this discounting of the liabilities includes a risk margin, as in the IASB definition of fair value, often varies.

Some believe that estimates of the change in the economic value of assets and liabilities represent a more meaningful measure of the gain or loss in a given period. But there are limitations of this measure:

- *Franchise Value* — To accurately reflect the change in value for a firm, changes in the value of its future profits must also be taken into account. This *franchise value* can be a significant source of value for firms (well in excess of the value of the assets and liabilities on its balance sheet) and changes in this value will clearly impact total shareholder returns.

- *GAAP Reconciliation* — The use of economic profit as an income measure also complicates reconciliation to GAAP income or other more familiar measures of profitability. This reconciliation issue is often important in practice because management may have more difficulty interpreting income measures that deviate significantly from commonly used measures.

- *External Communication* — If the economic profit measures are not disclosed to external parties such as investors, regulators or rating agencies, management may have more difficulty communicating the basis for their decisions. These external parties may only have access to GAAP and statutory financial statements and they may be unable to reproduce internally generated economic profit estimates.

Alternative Capital Measures for RAROC

There are numerous ways to measure the capital required for a given firm or for specific business units within the firm. Some of these capital measures are risk-adjusted and some are not.

Two measures that are not risk-adjusted include:

- *Actual Committed Capital* — This is the actual cash capital provided to the company by its shareholders and used to generate income for the firm and its respective business units. This is typically an accounting book value equal to contributed capital plus retained earnings and can be based on GAAP, Statutory or IASB accounting conventions.

- *Market Value of Equity* — The committed capital measure described above could be adjusted to reflect market values of the assets and liabilities, though this will still reflect only the value of the net assets on the balance sheet. An alternative is to actually use the market value of the firm's equity, which will generally be larger than the committed capital because of the inclusion of the franchise value of the firm.

Four measures that explicitly reflect risk-adjustments, to varying degrees, include:

- *Regulatory Required Capital* — This is the capital required to satisfy minimum regulatory requirements. This is typically determined by explicit application of the appropriate regulatory capital requirement model.

- *Rating Agency Required Capital* — This is the capital required to achieve a stated credit rating from one or more credit rating agencies (S&P, A.M. Best, Moody's or Fitch). This is usually determined by explicit application of the respective credit rating agencies' capital models and by reference to the standards each rating agency has established for capital levels required to achieve specific ratings.

- *Economic Capital* — This term is commonly used but often defined differently, which leads to unnecessary confusion. In its most general sense, economic capital could be defined as *the capital required to ensure a specified probability (level of confidence) that the firm can achieve a specified objective over a given time horizon.* The objective that the risk capital is intended to achieve can vary based on the circumstances and can vary depending upon whether the focus is on the policyholder, debtholder or shareholder perspectives.

i. Solvency Objective — The most common approach used by rating agencies and regulators could be referred to as a solvency objective. A solvency objective focuses on holding sufficient capital today to ensure that the firm can meet its existing obligations to policyholders (and perhaps debtholders as well). This approach clearly reflects a policyholder or debtholder perspective.

ii. Capital Adequacy Objective — An alternative approach is to use what could be referred to as a capital adequacy objective. This objective focuses on holding sufficient capital to ensure that the firm can continue to pay dividends, support premium growth in line with long-term business plans or maintain a certain degree of financial strength over an extended horizon so as to maximize the franchise value of the firm.

These two approaches can lead to substantially different indications of the capital required for the firm or any individual business. The solvency perspective is currently quite commonly used, so for convenience this perspective will be adopted throughout this paper. When using this definition of economic capital, the focus is typically on ensuring that there are sufficient financial resources (in cash and marketable securities) to satisfy policyholder (and debtholder) obligations. However, there will necessarily be a somewhat arbitrary separation of the total financial resources into a portion that represents a liability and a portion that represents capital. This separation will usually follow applicable accounting conventions, but can lead to meaningful differences in practice.

For instance, some practitioners define economic capital as the difference between the total financial resources needed less the undiscounted value of the (expected) liability. This is consistent with how the firm's resources would be classified under U.S. statutory accounting. Others prefer to define the economic capital as the amount that the total financial resources needed exceeds the discounted value of the (expected) liability. Still others might choose to incorporate a risk margin in the liability and treat the economic capital as the amount by which the total financial resources needed exceeds the fair value of the liability.

Any of these approaches could be used, so long as they are used consistently across different risks.

- *Risk Capital* — The range of different interpretations of the term economic capital is worrisome and can lead to a variety of inconsistent adjustments in practice. For instance, the choices described above all define economic capital as the portion in excess of the discounted expected liability, the undiscounted expected liability or the fair value of the liability under the assumption that funding for these amounts are already accounted for in the firm's financial statements. This is not the case for all risks — some could not be reflected at all on the balance sheet, in which case the economic capital has to account for all of the potential liabilities, while others could be funded by an amount well in excess of the discounted value, undiscounted value or fair value of the expected liability.

To avoid confusion in this paper, a closely related measure referred to here as *risk capital* will be used instead of any of the definitions of economic capital. Risk capital is defined as the amount of capital that must be contributed by the shareholders of the firm in order to absorb the risk that liabilities will exceed the funds already provided for in either the loss reserves or in the policyholder premiums. Under this definition, any conservatism in the loss reserves or any risk margins included in the premiums will reduce the amount of risk capital that must be provided by shareholders.

Notice that in the absence of a risk margin included in the premiums or the reserves, the risk capital and the economic capital may be identical. As a result, for many of the applications discussed later in this paper either amount could be used. However, for some of the main applications that involve evaluating specific business unit results or pricing for new business, the use of risk capital will more fairly account for the risk from the shareholder's perspective.

As a result, the term risk capital will be used here, even in instances where it is equivalent to the common definition of economic capital.

Risk Measures

The paper reviews various risk measures, including Value at Risk (percentile risk measure), Conditional Tail Expectation (CTE) and EPD Ratios, all of which are discussed in other readings on the syllabus and so will not be repeated here in these notes.

Aggregate Risk Profile

One approach to calculating a RAROC measure relies on an aggregate risk profile for the firm, which aggregates market, credit, insurance risks and perhaps other risks as well. For market and credit risks, techniques similar to those described in Hull's VaR and Credit Risk chapters are used. For the insurance risks, which include loss reserve risk, underwriting risk for prospective business to be written and property catastrophe risk, a variety of methods are used. Refer to the paper for detailed discussions of these models. Other risks that may be included, such as strategic risks or operational risks, are substantially more difficult to model and are ignored throughout the rest of the paper.

Various approaches are discussed for aggregating these risks into a single distribution of the aggregate risk. The main approach used later in the paper is to simulate an aggregate distribution using copulas (see Hull for an introduction to copulas and the paper for some additional details). But other approaches include closed form analytical calculations or approximations of the aggregate distribution.

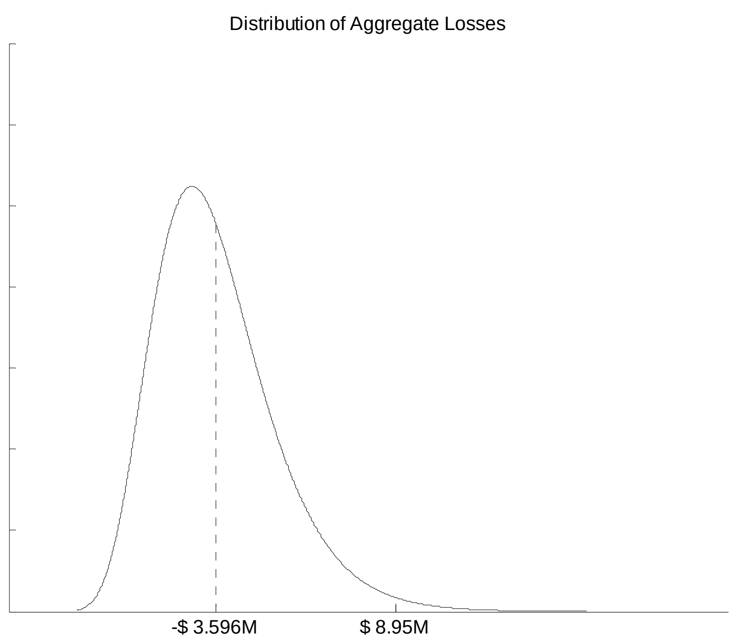
A final, but very different approach mentioned, is to simply aggregate the risk measures rather than to aggregate the risk distributions. This approach is commonly used (see Butsic's EPD paper) but has significant limitations.

Capital Allocation

The paper provides a lot of details regarding capital allocation to various risk categories, which builds on the material discussed in the Cummins paper. A hypothetical insurance company, Sample Insurance Company (SIC), is used to estimate risk distributions for market risk, loss reserve risk for one existing line of business and underwriting risk for two lines of business. All other risk categories (credit, property catastrophe, etc.) are ignored for the numerical examples.

The aggregate risk distribution for SIC is then obtained by simulating from each of the stand-alone risk distributions and using a normal copula to account for the desired correlation/dependency among the risk types. For presentation purposes, a lognormal distribution was fit to the empirical aggregate distribution using the method of moments. The resulting distribution is shown in Figure 1.

FIGURE 1. Distribution of Aggregate Losses — Sample Insurance Company



For the purposes of this distribution, the amounts shown represent the potential loss, with profits depicted as negative amounts and losses as positive amounts. Because both assets and liabilities are included in this calculation, the aggregate loss includes the losses (or profits) in the investment portfolio and the insurance claim costs and expenses in excess of the premiums.

The company was assumed to have \$19.6 million of invested assets initially, collected \$12.8 million of premium for the two lines of business and paid 5% of that premium in expenses. These amounts are available to pay claims but are assumed to be contributed by the policyholders, to distinguish them from the risk capital that must be contributed by shareholders.

From this aggregate risk distribution, the 99th percentile risk measure is \$8.95 million. If this amount of risk capital were contributed to the firm the probability of having insufficient assets to pay all of the claims fully would be 1% (ignoring the effect of market risk beyond the one-year horizon for simplicity).

Next, several capital allocation approaches are demonstrated using the SIC numerical example described above. In each case, the \$8.95 million of risk capital will be allocated to some or all of the major risk sources. The methods discussed include the following:

- *Proportional Allocation Based on a Risk Measure* — This method simply calculates stand-alone risk measures for each risk source (market risk, reserve risk, Line A underwriting risk, Line B underwriting risk) and then allocates the total risk capital in proportion to the separate risk measures.

- *Incremental Allocation* — This method determines the impact that each risk source has on the aggregate risk measure and allocates the total risk capital in proportion to these incremental amounts.

- *Marginal Allocation (Myers-Read Method)* — This method determines the impact of a small change in the risk exposure for each risk source (e.g. amount of assets, amount of reserves, premium volume) and allocates the total risk capital in proportion to these marginal amounts. One particular method that will be demonstrated is the Myers-Read method.

- *Co-Measures Approach (Kreps, Ruhm-Mango)* — This method determines the contribution each risk source has to the aggregate risk measure. The method that was independently developed by Kreps and by Ruhm and Mango will be demonstrated.

The paper then shows numerical examples of each of these methods, using the model and assumptions described earlier for Sample Insurance Company.

Proportional Allocation Based on a Risk Measure

Using any selected risk measure, such as a percentile risk measure (VaR) or the CTE, each unit's proportional risk measure to the sum of all the risk measures is applied to the total capital. For example, if the stand-alone 99th percentile risk measure (referred to here as the 99% VaR) is used for each risk source, then the allocation shown in Table 1 on the following page is obtained.

Because the 99% VaR risk measure was used to determine the aggregate capital, it seems reasonable to use the same risk measure to perform the allocation. However, some practitioners choose to use a different risk measure as the basis for allocating risk than is used to measure the aggregate risk capital.

Table 1. Capital Allocation — Proportional to 99% VaR

	99.00% VaR	% of Total	Allocated Capital
Market Risk	1,183,461	8%	742,665
Reserve Risk	4,440,453	31%	2,786,545
Line A UW Risk	3,243,793	23%	2,035,598
Line B UW Risk	5,394,016	38%	3,384,941
Total	14,261,723		8,949,750

For instance, if the 99.97% VaR is used to allocate risk capital but the same total amount of risk capital from the previous example is allocated, the results shown in Table 2 on the next page are obtained.

Table 2. Capital Allocation — Proportional to 99.97% VaR

	99.97% VaR	% of Total	Allocated Capital
Market Risk	2,500,702	10%	851,813
Reserve Risk	8,035,878	31%	2,737,259
Line A UW Risk	5,666,239	22%	1,930,089
Line B UW Risk	10,071,313	38%	3,430,588
Total	26,274,131		8,949,750

Similarly, if the 99% CTE were used as the risk measure, the results in Table 3 would be obtained.

Table 3. Capital Allocation — Proportional to 99% CTE

	99.00% CTE	% of Total	Allocated Capital
Market Risk	1,593,170	9%	799,365
Reserve Risk	5,441,265	31%	2,730,126
Line A UW Risk	3,922,399	22%	1,968,043
Line B UW Risk	6,880,426	39%	3,452,217
Total	17,837,260		8,949,750

Notice that in all three cases here, the allocations are quite similar. In other applications, particularly those that include highly skewed risks such as property-catastrophe risk, this will not always be the case. In addition, there are many instances where it may be appropriate to use risk measures that are not tail based. In these instances, the differences that result could be more significant.

For example, Table 4 on the next page and Table 5 on the facing page show the allocations that would result from using the 80% VaR or the 80% CTE risk measures.

Notice that in the 80% VaR allocation, the market risk allocation is negative. This reflects the fact that at the 80th percentile of the market risk distribution, the market returns are positive

TABLE 4. Capital Allocation — Proportional to 80% VaR

	80.00% VaR	% of Total	Allocated Capital
Market Risk	-586,016	-35%	-3,132,546
Reserve Risk	335,121	20%	1,791,389
Line A UW Risk	756,744	45%	4,045,176
Line B UW Risk	1,168,409	70%	6,245,731
Total	1,674,258		8,949,750

TABLE 5. Capital Allocation — Proportional to 80% CTE

	80.00% CTE	% of Total	Allocated Capital
Market Risk	80,957	1%	117,902
Reserve Risk	1,809,817	29%	2,635,718
Line A UW Risk	1,622,380	26%	2,362,745
Line B UW Risk	2,632,196	43%	3,833,385
Total	6,145,350		8,949,750

and reduce the aggregate capital needs. The impact of this is offset significantly in the 80% CTE allocation because the entire tail of the market risk distribution is taken into account and therefore the scenarios in which the market returns are negative impact the overall capital allocation to market risk.

Incremental Allocation

Under this approach, an aggregate risk measure is selected and calculated for the aggregate risk distribution. Then, the same risk measure is recalculated after removing one of the business units. The difference in the capital requirement with and without the selected business unit then represents the incremental capital requirement for the business unit.

Using the incremental capital requirements for each business unit, the firm's capital can then be allocated to each unit in proportion to its respective incremental capital requirements. This is demonstrated in Table 6.

TABLE 6. Capital Allocation — Incremental Based on 99% VaR

	Total 99.00% VaR	All-Other 99.00% VaR	Incremental 99.00% VaR	% of Total	Allocated Capital
Market Risk	8,949,750	8,661,043	288,707	3%	241,168
Reserve Risk	8,949,750	5,510,089	3,439,661	32%	2,873,285
Line A UW Risk	8,949,750	5,869,650	3,080,099	29%	2,572,929
Line B UW Risk	8,949,750	5,044,312	3,905,437	36%	3,262,367
Total			10,713,904		8,949,750

An important characteristic of this allocation method is that the incremental amounts do not add up to the total capital, even though the same risk measure was used. This is a characteristic that some practitioners find troublesome and there is disagreement over whether the excess amount should be allocated.

Marginal Allocation

The incremental allocation eliminates an entire business unit to determine its capital requirements. Instead, one could eliminate one dollar of revenue or one dollar of expected loss from each unit sequentially and use the change in the firm's total capital requirement as an estimate of the marginal capital requirement for the unit.

Applying this marginal requirement to the total revenue or total expected losses for the business unit provides an alternative measure of the capital needed for the unit. This can then be allocated in the same manner as described above for the incremental allocation method.

This approach typically results in a more appropriate result, however it may be impractical in certain circumstances where not all risk sources can be represented relative to revenue or expected loss or their marginal impacts easily determined.

Myers-Read Method

This is a specific type of marginal allocation method, but its basis is somewhat different than those described above. Because an insurance company's total potential losses almost always exceed its assets, its owners have an option to default on the firm's obligations and put the claims (or some portion thereof) back to the policyholders. The value of this put option will decline as the amount of capital held increases for the same exposures. The Myers-Read method allocates capital so as to equalize the marginal impact that each business unit has on the value of this put option.

To apply this method, the value of the default option is calculated based on the firm's current capital and its current exposures. The exposure for a given business unit is then increased and the capital needed to maintain the same value of the firm's aggregate default option is determined. This capital then represents a marginal requirement per unit of expected loss for each unit that can be applied to the unit's expected losses.

The results of this method are demonstrated in Table 7 on the facing page (see Appendix B in the paper for the technical details). For this example, the target EPD Ratio has been set arbitrarily to 0.186% so that the resulting aggregate risk capital is identical to the 99% VaR risk measure used in the other allocation method examples. In addition, the methodology takes into account the market risk in the invested assets, though it does not allocate capital to the market risk component. All capital is allocated to the lines of business for which there are liabilities, since it is only the need to pay liabilities that gives rise to the need to hold capital.

TABLE 7. Capital Allocation — Myers-Read (0.186% EPD Ratio)

	Capital to Loss Ratio	Expected Claims	Capital	% of Total	Allocated Capital
Reserve Risk	21.78%	18,091,233	3,939,466	44%	3,939,466
Line A UW Risk	33.92%	5,860,732	1,988,079	22%	1,988,079
Line B UW Risk	51.57%	5,860,732	3,022,205	34%	3,022,205
Total			8,949,750		8,949,750

This particular method has become popular because it produces additive capital requirements that sum to the total capital requirement for the firm when the same risk measure is used. Three points are worth noting with respect to this method:

i. *Developed to Allocate Frictional Costs* — The method was not developed as a means for determining risk-adjusted capital requirements; it was developed as a means to allocate the frictional costs of capital to various businesses. While it may be used for the former purpose, it is not necessarily more appropriate for this purpose than the other methods discussed.

ii. *Relies on Default Option Framework* — Because this method requires the valuation of the default option, its application may require substantially more quantitative resources compared to other methods, except in certain limited circumstances.

iii. *Mathematical Limitations* — Significant mathematical challenges have been raised that suggest that the Myers-Read method is not appropriate for most insurance applications. The method assumes that risk exposure in a business unit can be increased or decreased without impacting the shape of the loss distribution, a property referred to as homogeneity. Except when risk can be increased or decreased through changes in quota share percentages, insurance loss distributions will not exhibit homogeneity when adding or removing policies from the firm's mix of business.

Co-Measures Approach

This approach establishes the firmwide capital requirement using a particular conditional risk measure, such as VaR or CTE, and then calculates the Co-Measure for each business unit by calculating the comparable risk measure for the unit subject to the condition applied to the entire firm.

For example, consider the case where the risk measure selected is the CTE. The firmwide CTE is the average value of the losses given that the losses for the firm exceed the Value at Risk at a chosen percentile. To determine the Co-CTE for a given business unit, simply calculate the average losses for each business subject to the same firmwide condition that the total losses for the firm exceeds the chosen percentile.

This is very easy to implement in a simulation context. For example, the four key risk components for the SIC example were simulated using a normal copula method and the aggregate loss was determined for each of 50,000 simulation scenarios. The results were sorted in descending order based on the total loss and the worst 1% of the scenarios (the top 500 scenarios) were identified, as shown in Table 8.

Table 8. Co-CTE Calculations

Sorted Scenario	Market	Reserves	Line A	Line B	Total
1	779,323	12,180,298	3,188,429	4,994,583	21,142,632
2	494,425	8,169,822	3,734,913	8,695,665	21,094,825
3	-3,407,081	13,140,377	7,607,985	788,471	18,129,751
4	-779,922	2,587,705	5,675,660	10,386,216	17,869,658
5	-1,311,004	-1,203,142	3,238,333	16,924,158	17,648,345
6	-1,392,828	5,488,457	6,646,703	6,799,820	17,542,152
7	-255,475	4,812,487	4,018,249	7,904,885	16,480,145
8	-10,210	6,710,721	2,273,968	7,472,474	16,446,953
9	-1,896,169	4,433,724	1,652,542	12,169,231	16,359,328
10	758,494	3,132,459	2,330,630	10,003,805	16,225,388
11	-1,291,494	8,133,807	5,475,393	3,899,206	16,216,912
12	1,523,399	8,164,027	1,320,562	4,996,263	16,004,250
13	-1,507,026	8,701,922	4,941,913	3,358,494	15,495,303
14	-418,192	-390,473	1,172,596	15,112,222	15,476,153
15	348,569	4,904,846	4,173,982	6,001,026	15,428,423
:	:	:	:	:	:
490	-470,761	3,622,090	-148,615	4,519,262	7,521,976
491	-980,559	3,630,412	1,980,834	2,889,533	7,520,220
492	-2,921,510	2,906,628	-200,015	7,730,833	7,515,936
493	-1,179,044	3,552,559	2,343,631	2,794,807	7,511,953
494	-2,744,202	2,173,409	4,717,356	3,364,141	7,510,703
495	127,947	1,318,389	4,749,312	1,308,659	7,504,307
496	42,016	1,663,231	1,653,643	4,143,005	7,501,894
497	-1,062,298	2,170,695	6,366,285	27,183	7,501,865
498	-901,735	4,579,393	-124,816	3,947,145	7,499,986
499	-2,782,565	972,163	1,896,786	7,411,779	7,498,163
500	-2,959,845	6,146,281	863,894	3,441,193	7,491,523
Co-CTE	**-908,399**	**3,715,533**	**2,279,319**	**4,549,138**	**9,635,591**

The overall 99% CTE is simply the average total loss for the 500 worst scenarios, or $9.635 million. For each of these specific scenarios, the four main risk components make a different contribution to the total loss. For example, in Scenario 1, 58% of the total loss came from the reserve risk, 24% came from Line B's underwriting risk, 15% came from Line A's underwriting risk and 3% came from the market risk.

Note that, on average over these 500 scenarios, the market risk component actually reduced the total loss (due to profits in the investment portfolio rather than losses). Taking an average for each of these risk components, not across each of their own respective worst 1% of outcomes but rather across these specific 500 scenarios that represent the worst 1% of the total outcomes, the Co-CTE's are calculated as shown in the bottom row of the table. These reflect the average contribution each makes to the total losses.

TABLE 9. Capital Allocation — Proportional to 99% Co-CTE

	99.00% Co-CTE	% of Total	Allocated Capital
Market Risk	-908,399	-9%	-843,742
Reserve Risk	3,715,533	39%	3,451,069
Line A UW Risk	2,279,319	24%	2,117,082
Line B UW Risk	4,549,138	47%	4,225,340
Total	9,635,591		8,949,750

As shown in this table, on average the reserve risk contributes 39% of the total losses, Line A's underwriting risk contributes 24% of the total losses and Line B's underwriting risk contributes 47% of the total losses.

In addition, the Co-CTE's *add-up* to the total CTE as shown in the bottom row of the scenario summary. But to remain consistent with the other allocation examples and to highlight the ability to separate the allocation method from the amount allocated, the final allocation in the last column uses the Co-CTE allocation percentages applied to the 99th percentile risk measure (99% VaR) total risk capital figure used earlier.

Applications

The paper discusses five uses of the calculations described above. The two most relevant for the exam are shown in bold in the list below and are described more fully in the sections that follow.

- Assessing Capital Adequacy
- Setting Risk Management Priorities
- Evaluating Alternative Risk Management Strategies
- **Risk-Adjusted Performance Measurement**
- **Insurance Policy Pricing**

RAROC Application

It is often desirable to evaluate actual, *ex post*, performance of different business units. Traditional measures of financial performance for insurers, such as historical loss ratios, can provide misleading indications of relative results for two business units with different levels of risk. For instance, if a business unit with a high degree of risk were to have a lower loss ratio than a business unit with a low amount of risk, the loss ratios alone may not properly identify which of the two business units performed better. The use of a risk-adjusted performance metric such as RAROC may allow these business units to be more fairly compared. The explicit risk-adjustment may also be an improvement over judgmental premium to surplus ratios.

As an example of this process, consider the Sample Insurance Company presented in Section 4 of the paper and summarized above. Rather than rely upon the expected loss ratios, hypothetical values for the actual loss ratios realized over the year will be used. For this example, the actual loss ratios will be assumed to equal 92% for Line A and 86% for Line B.

Based solely on the loss ratios, it is natural to assume that Line B performed better. Calculation of an economic profit could also be used to show that Line B had a larger present value profit. For example, assuming that the actual market returns were 5%, then each line of business would have had economic profit at the end of the year as shown in Table 10.

TABLE 10. Calculation of Actual Economic Profit

		Line A	Line B	Calculations
(1)	Premium	6,400,000	6,400,000	Actual
(2)	Expense Ratio	5.00%	5.00%	Actual
(3)	Expenses	320,000	320,000	(3) = (1) * (2)
(4)	Investment Return	5.00%	5.00%	Actual
(5)	Investment Income	304,000	304,000	(5) = (4) * [(1) - (3)]
(6)	Discounted Loss Ratio	92.00%	86.00%	Actual
(7)	Discounted Claim Costs	5,888,000	5,504,000	(7) = (6) * (1)
(8)	**Economic Profit**	**496,000**	**880,000**	(8) = (1) - (3) + (5) - (7)

As shown in Section 4 of the paper, Line B exposed the firm to substantially more risk than Line A and its profit per dollar of risk capital was actually lower.

For instance, Table 11 shows the RAROC for these business units if the 99% Co-CTE allocation method were used.

TABLE 11. Comparison of RAROC — Using 99% Co-CTE Allocation

	Economic Profit	Allocated Capital	RAROC
Line A	496,000	2,117,082	23.4%
Line B	880,000	4,225,340	20.8%

By rescaling the profit by the allocated capital for the underwriting risk, the risk-adjusted profitability measure shows that despite the lower loss ratio and higher economic profit, Line B required far more capital to support its operations and as a result did not outperform Line A.

This use of RAROC to better inform the assessment of performance shows that it is possible to take risk into consideration in a relatively simple manner. However, there are a variety of allocation methods that could be used. If proportional allocation based on the 99th percentile (99% VaR) risk measure were used instead, the results would be as shown in Table 12.

TABLE 12. Comparison of RAROC — Using Proportional 99% VaR Allocation

	Economic Profit	Allocated Capital	RAROC
Line A	496,000	2,035,598	24.4%
Line B	880,000	3,384,941	26.0%

This comparison shows that RAROC, despite its appeal as a means to risk-adjust performance metrics, does not necessarily produce unambiguously superior performance measures. Depending upon the method used for the allocation, the RAROC for Line B could be either lower than or higher than the RAROC for Line A. These results are highly sensitive to a variety of implicit and explicit assumptions that can materially impact the allocation of capital to specific business units.

Pricing Application

A natural extension of the RAROC analysis just demonstrated, which focused on a relative comparison of two business units, is to use RAROC directly in insurance policy pricing. The rationale would be to set the price such that the expected RAROC is above a specified target rate.

Suppose, for instance, that an acceptable RAROC target of 15% is assumed. The premium that should be charged such that Line B's expected RAROC was equal to at least 15% would then be easy to determine. One approach, albeit overly simplified and somewhat naive, is to simply choose one of the many capital allocation methods and then solve for the additional risk margin, which will be denoted by π here, such that the RAROC equals the target rate of 15%.

For the sake of a numerical example, consider the allocation of risk capital to Line B using the Co-CTE allocation method. Based on the existing assumptions regarding Line B's expected loss ratio rather than the actual loss ratio used in the previous example, this produces the expected economic profit and expected RAROC for Line B shown in Table 13 on the next page and Table 14 on the following page, respectively.

With no additional risk margin, the RAROC is below the target rate. The following equation can be used to solve for the additional risk margin, π, that produces the target rate of 15% (under

TABLE 13. Expected Economic Profit — Line B

		Line B	Calculations
(1)	Premium	6,400,000	Expected
(2)	Expense Ratio	5.00%	Expected
(3)	Expenses	320,000	(3) = (1) * (2)
(4)	Investment Return	5.00%	Expected
(5)	Investment Income	304,000	(5) = (4) * [(1) - (3)]
(6)	Discounted Loss Ratio	91.60%	Expected
(7)	Discounted Claim Costs	5,862,400	(7) = (6) * (1)
(8)	**Expected Economic Profit**	**521,600**	(8) = (1) - (3) + (5) - (7)

TABLE 14. Expected RAROC — Using 99% Co-CTE Allocation

	Expected Economic Profit	Allocated Capital	Expected RAROC
Line B	521,600	4,225,340	12.3%

the simplifying assumption that the increased risk margin does not impact expenses).

$$\begin{aligned} P &= \text{Original Premium} \\ E &= \text{Expenses} \\ i &= \text{Expected Investment Income} \\ PV(L) &= \text{PV of Expected Claims} \end{aligned}$$

$$\begin{aligned} \text{RAROC} &= \frac{[P + \pi - E](1 + i) - PV(L)}{\text{Allocated Risk Capital}} \\ &= \frac{[6{,}400{,}000 + \pi - 320{,}000](1 + 5\%) - 5{,}862{,}400}{4{,}225{,}340} \\ &= 15\% \\ \pi &= 106{,}858 \end{aligned}$$

Economic Value Added

This solution can also be derived using what is often referred to as an Economic Value Added or EVA approach.

If the $4,255,340 is treated as the required capital to write Line A and the 15% RAROC target is the per unit cost of capital, then the total dollar cost of the capital is:

$$15\%(\$4{,}255{,}340) = \$633{,}801$$

This is the amount of expected economic profit that would have to be incorporated into the premium.

Since the original premium already accounted for $521,600 of this expected profit, only $112,201 of additional profit would be required at the end of the year. Discounting that amount to reflect the fact that it is collected up-front, $106,858 of additional risk margin would have to be incorporated into the up-front premium to meet the RAROC target rate.

These calculations are shown in Table 15.

TABLE 15. Calculation of Additional Risk Margin Required

		Amount	Calculations
(1)	Allocated Risk Capital	4,225,340	Based on Co-CTE Allocation
(2)	Target RAROC	15.0%	Assumed
(3)	Required Economic Profit	633,801	(3) = (1) * (2)
(4)	Current Economic Profit	521,600	Based on Assumptions
(5)	Shortfall	112,201	(5) = (3) - (4)
(6)	Expected Investment Income	5.00%	Based on Assumptions
(7)	Additional Risk Margin Required	106,858	(7) = (5)/[1 + (6)]

Notice that in this calculation the additional risk margin is assumed to earn the same expected rate of investment income as the net premiums. An argument could be made that the additional risk margin should be assumed to be invested in risk-free assets only, to avoid the need to calculate the additional risk capital that investing these funds in risky assets might produce. But the impact of this is likely to be small and can usually be ignored.

Additional Considerations for Pricing Application

Using RAROC for pricing, as in the above example, is appealing because the steps are logical and easy to explain. However, some subtle complications can arise in practice that are not as obvious in this example due to some of the simplifications made. The consequences of three specific simplifications of importance to pricing applications are discussed, with additional complications relevant to all applications discussed in Section 6 of the paper.

The three specific issues relevant to pricing applications are:

i. Multi-Period Capital Commitment

ii. Cost of Capital (Target RAROC Rate)

iii. Investment Income on Allocated Capital

Multi-Period Capital Commitment

Up until now, the allocated risk capital was assumed to be exposed to risk for only a single period. This allowed the discussion of RAROC to be somewhat simplified and did not impact any of the conclusions drawn from the previous examples, in part because both Line A and Line B had the same claim payment patterns and the comparisons were made relative to each other rather than on an absolute basis.

But in the context of policy pricing, it is important to recognize that the initial capital required to write the policy does not fully reflect the total capital costs. The risk will not be fully resolved in a single period and so some risk capital will be needed in subsequent periods as well, perhaps until the final claims are paid. It is easy to see how one might account for this in practice. One common approach is to assume an average pattern for the release of the risk capital and then use that pattern either to adjust the RAROC ratio or to modify the target rate.

To see how these adjustments could be made, consider an assumed claim payment pattern for Line A as shown in Table 16 (chosen arbitrarily for simplicity). Further, assume that the risk

TABLE 16. Claim Payment Pattern — Line A

Year	% Paid
1	50%
2	30%
3	15%
4	5%

capital will be released, on average, at the same rate as the claims are paid. In reality, under some scenarios more capital will be released, perhaps faster or slower than this pattern, and under some scenarios more capital may even be committed to support this line of business. But given the assumed release pattern for the allocated risk capital, the cost per unit of risk capital (15% for the sake of this example) can be applied to the outstanding risk capital each period and the aggregate cost of risk capital over the life of the exposures can be determined. The calculations are shown in Table 17 on the next page.

Using the same approach as before, the risk margin can be determined to ensure that it is sufficient to compensate the capital providers for the total cost of the risk capital over its life.

Cost of Risk Capital

In the above discussion, a constant 15% cost of risk capital was assumed, without explanation or justification. It is worth exploring, particularly in the context of insurance policy pricing, how this cost of risk capital should be determined in practice.

TABLE 17. Aggregate Cost of Risk Capital — Multi-Period Release of Risk Capital

Year	% Paid	Beginning Risk Capital	Cost of Risk Capital	PV Cost of Risk Capital	Risk Capital Released	Ending Risk Capital
1	50%	4,225,340	633,801	603,620	2,112,670	2,112,670
2	30%	2,112,670	316,901	287,438	1,267,602	845,068
3	15%	845,068	126,760	109,500	633,801	211,267
4	5%	211,267	31,690	26,071	211,267	0
			1,109,152	1,026,630		

Although the RAROC measure is intuitively appealing, it is more ad hoc than many practitioners often recognize. Because it is referred to as a return on capital, it is quite common for practitioners to assume that standard *rate of return* benchmarks, such as those derived from models such as Capital Asset Pricing Model (CAPM) or the Fama-French 3-Factor Model, are applicable. In reality, the rate used for the cost of risk capital must take into account the specific way in which the RAROC metric is defined.

The most significant issues include the following:

- *Definition of Risk* — Numerous textbook discussions of RAROC suggest using risk-adjusted return models such as CAPM to establish the cost of risk capital and to assess whether or not the RAROC exceeds this value. Despite the fact that both RAROC and CAPM produce "risk-adjusted returns", the risk adjustment in RAROC reflects a different definition of risk than is used in these theoretical models.

  Models such as CAPM measure the systematic risk associated with an investment, which accounts for the marginal contribution the investment adds to an existing portfolio of diversified investments. RAROC, even for the total firm, incorporates an entirely different measure of risk based on the relationship between a cash flow's expected value and certain values in the tail of its probability distribution.

- *Leverage* — RAROC is artificially leveraged because the denominator reflects neither the total market value of the invested capital (as is assumed in the theoretical return models) nor the firm's actual capital that could be exposed to loss (the committed capital).

  If the firm's shareholders desire a given target rate of return on their investment, the dollar value of their target income will depend on the total market value of the firm's equity. This will almost always exceed the value of the firm's book equity (the difference being attributed to their franchise value), though under certain assumptions regarding the stability of the firm's market to book value multiple, the rate of return on market value and the rate of return on book value may be equal.

  Nevertheless, earning this rate of return solely on the firm's risk capital will not necessarily be sufficient to satisfy the income expectations of the shareholders. Using

> this lower base in the denominator of RAROC artificially inflates the rate of return on capital, with only a modest offset due to the fact that the numerator also ignores a component of income based on changes in the franchise value of the firm.
>
> When RAROC is measured for distinct business units within the firm, the capital allocated to those business units will depend upon the degree to which diversification effects are reflected in the amount allocated, the risk measure used and other somewhat arbitrary decisions. The business unit losses are not literally limited to the amount of risk capital allocated to it, so this leverage effect on the RAROC is even more artificial.

Taking these considerations into account is a bigger challenge than is often recognized and entirely satisfactory methods for calibrating the cost of risk capital do not exist.

One acceptable compromise is to recognize that models such as CAPM or the Fama-French 3-Factor Model are reasonable means to quantify shareholders' target return on the firm's total capital (e.g. GAAP book value). Under a conservative assumption that only the total risk capital is at risk, the CAPM return can be adjusted upwards by the ratio of the firm's total capital to the firm's risk capital. Alternatively, rather than using the (arbitrary) risk capital in the RAROC calculation, the firm's total capital could be used along with the allocation methods discussed here. In either case, this allows the pricing model to reflect the aggregate compensation required by the shareholders for assuming systematic risk and then allocates this total amount to different business units (or policies) in a manner that reflects the relative risk of each.

This approach does not account for the differential degrees of leverage in each business unit. This is because after taking into account diversification benefits, it is quite difficult to quantify how much additional leverage has been introduced into the calculation.

This approach also does not address the potential for different business units to have different degrees of systematic risk. Theoretically this should be easy to deal with, though in practice adjustments to reflect differing degrees of systematic risk across segments of the total firm are quite difficult to make because of the limited ability to reliably quantify these differences.

Many alternative methods for quantifying the cost of risk capital have been proposed. For example, Feldblum suggests incorporating the frictional costs of holding capital, such as those that result from the double taxation of investment income. Venter points out though that this is incomplete because it doesn't address the compensation required for assuming the risk that would reasonably need to be included even in the absence of corporate taxes (e.g. for Bermuda-based reinsurance firms that are not subject to corporate income taxes).

Yet another approach has been suggested by Mango. Mango notes that while it is common to refer to the allocation of capital, it is really just an allocation of underwriting capacity and therefore the policy or business unit must earn adequate profits to pay for this capacity. In addition, each policy or business unit also is given the ability to call upon the capital not explicitly allocated to it, if needed to pay claims, and therefore must also earn adequate profits to compensate the firm for the value of this capital call. These costs could be combined and reflected as a cost per unit of allocated risk capital.

Investment Income on Allocated Capital

In the simplified example shown above, it was assumed that the target return on the allocated risk capital was 15%. How this target return is calculated depends on how the economic profit is defined.

The definition of economic profit used in the example above did not include the investment income that can be expected to be earned on the allocated risk capital itself. As a result, the 15% target return also excludes the investment rate assumed to be earned on the allocated risk capital. The target return is technically an excess return over the investment rate.

Alternatively, the investment income expected to be earned on the allocated risk capital could be included in the calculation of economic profit. In this case, the target return should include the investment rate assumed to be earned on the allocated risk capital. In a single period context, the two approaches lead to the same risk margin. However, when risk capital is required over multiple periods, the approach used in the examples above is easier to apply.

Practice Questions

Question 1. List and briefly describe the key risk sources used to model a P&C insurer's aggregate risk profile.

Solution. The four broad categories are:

i. Market risk — Reflects potential loss in value of invested assets due to changes in interest rates, equity markets, foreign exchange rates, etc.

ii. Credit risk — Reflects potential loss in value due to counterparty default or changes in counterparty credit ratings.

iii. Insurance risk — Includes potential losses from loss reserves, prospective underwriting and property catastrophe exposure.

iv. Other — Other risks that may be included are strategic risks or operational risks.

Question 2. What methods are discussed for aggregating these different risks?

Solution. The main approach used later in the paper is to simulate an aggregate distribution using copulas (see Hull for an introduction to copulas and the paper for some additional details). But other approaches include closed form analytical calculations or approximations of the aggregate distribution.

A final, but very different approach mentioned, is to simply aggregate the risk measures rather than to aggregate the risk distributions. This approach is commonly used (see Butsic's EPD paper) but has significant limitations.

Question 3. List and briefly describe the four main methods used in the paper to allocate capital to different risk sources.

Solution. The four methods are:

- Proportional Allocation Based on a Risk Measure — This method simply calculates stand-alone risk measures for each risk source (market risk, reserve risk, Line A underwriting risk, Line B underwriting risk) and then allocates the total risk capital in proportion to the separate risk measures.

- Incremental Allocation — This method determines the impact that each risk source has on the aggregate risk measure and allocates the total risk capital in proportion to these incremental amounts.

- Marginal Allocation (Myers-Read Method) — This method determines the impact of a small change in the risk exposure for each risk source (e.g. amount of assets, amount of reserves, premium volume) and allocates the total risk capital in proportion to these marginal amounts. One particular method demonstrated in the paper is the Myers-Read method.

- Co-Measures Approach — This method determines the contribution each risk source has to the aggregate risk measure.

Question 4. In the paper, a numerical example was used to simulate the aggregate risk profile for a P&C insurer, producing 50,000 iterations. The 15 worst results, when sorted on the total loss amount, are shown in the following table:

Sorted Scenario	Market	Reserves	Line A	Line B	Total
1	779,323	12,180,298	3,188,429	4,994,583	21,142,632
2	494,425	8,169,822	3,734,913	8,695,665	21,094,825
3	-3,407,081	13,140,377	7,607,985	788,471	18,129,751
4	-779,922	2,587,705	5,675,660	10,386,216	17,869,658
5	-1,311,004	-1,203,142	3,238,333	16,924,158	17,648,345
6	-1,392,828	5,488,457	6,646,703	6,799,820	17,542,152
7	-255,475	4,812,487	4,018,249	7,904,885	16,480,145
8	-10,210	6,710,721	2,273,968	7,472,474	16,446,953
9	-1,896,169	4,433,724	1,652,542	12,169,231	16,359,328
10	758,494	3,132,459	2,330,630	10,003,805	16,225,388
11	-1,291,494	8,133,807	5,475,393	3,899,206	16,216,912
12	1,523,399	8,164,027	1,320,562	4,996,263	16,004,250
13	-1,507,026	8,701,922	4,941,913	3,358,494	15,495,303
14	-418,192	-390,473	1,172,596	15,112,222	15,476,153
15	348,569	4,904,846	4,173,982	6,001,026	15,428,423
Sum	-8,365,192	88,967,035	57,451,858	119,506,519	257,560,220
Average	-557,679	5,931,136	3,830,124	7,967,101	17,170,681

Determine the Co-CTE's for each component and the CTE for the total using the 99.97% CTE.

Solution. The 99.97% CTE is the average for all scenarios that are worse than the 99.97th percentile. Since there are 50,000 scenarios, we are interested in the 1 – 99.97% = 0.03% worst scenarios. That is, we are interested in the worst 15 scenarios.

To calculate the aggregate CTE, we simply calculate the average value of the worst 15 scenarios. Using the numbers in the last column, this is equal to $17,170,681. The Co-CTE's for each of the four components are determined the same way, using the same scenarios (sorted on the aggregate column, not each respective column).

The results for each component are shown below:

TABLE 18. 99.97% Co-CTE

	99.97% Co-CTE
Market Risk	-557,679
Reserve Risk	5,931,136
Line A UW Risk	3,830,124
Line B UW Risk	7,967,101
Total	17,170,681

Question 5. Suppose that for the previous question we wanted to define the aggregate risk capital using a different risk measure than the 99.97% CTE but still wanted to use the 99.97%

CTE to determine the allocation. Using $8,949,750 as the aggregate risk capital, how is this allocated to each component?

Solution. The allocation uses the Co-CTE's in proportion to their total, as shown below:

TABLE 19. Capital Allocation — Proportional to the 99.97% Co-CTE

	99.97% Co-CTE	% of Total	Allocated Capital
Market Risk	-557,679	-3%	-290,675
Reserve Risk	5,931,136	35%	3,091,443
Line A UW Risk	3,830,124	22%	1,996,348
Line B UW Risk	7,967,101	46%	4,152,634
Total	17,170,681		8,949,750

Question 6. The paper discusses the fact that a wide variety of definitions could be used for the income measurement in RAROC. How is the *economic profit* calculated in the paper for the purposes of calculating an *ex post* RAROC?

Solution. In the paper, the profit is measured at the end of one year. The premium, net of up-front expenses, is adjusted to reflect one period of investment income on the net premium and then the claims are subtracted on a discounted basis, with the discounting being done to the end of the year for all payments expected to be made beyond that time horizon.

Question 7. Use the capital allocation in Question 5 and the following economic profit calculations for Line A and Line B to determine the RAROC for each.

TABLE 20. Calculation of Actual Economic Profit

		Line A	Line B	Calculations
(1)	Premium	6,400,000	6,400,000	Actual
(2)	Expense Ratio	5.00%	5.00%	Actual
(3)	Expenses	320,000	320,000	(3) = (1) * (2)
(4)	Investment Return	5.00%	5.00%	Actual
(5)	Investment Income	304,000	304,000	(5) = (4) * [(1) - (3)]
(6)	Discounted Loss Ratio	92.00%	86.00%	Actual
(7)	Discounted Claim Costs	5,888,000	5,504,000	(7) = (6) * (1)
(8)	Economic Profit	496,000	880,000	(8) = (1) - (3) + (5) - (7)

Solution. Using the results provided, the RAROCs are:

TABLE 21. RAROC Using 99.97% Co-CTE Allocation of Capital

	Economic Profit	Allocated Capital	RAROC
Line A	496,000	1,996,348	24.8%
Line B	880,000	4,152,634	21.2%

Question 8. Suppose you have estimated the expected economic profit for a line of business and have determined how much capital to allocate or attribute to this line to reflect its risk. Your calculations are as shown below:

TABLE 22. Expected Economic Profit — Line B

		Line B	Calculations
(1)	Premium	6,400,000	Expected
(2)	Expense Ratio	5.00%	Expected
(3)	Expenses	320,000	(3) = (1) * (2)
(4)	Investment Return	5.00%	Expected
(5)	Investment Income	304,000	(5) = (4) * [(1) - (3)]
(6)	Discounted Loss Ratio	91.60%	Expected
(7)	Discounted Claim Costs	5,862,400	(7) = (6) * (1)
(8)	Expected Economic Profit	521,600	(8) = (1) - (3) + (5) - (7)

TABLE 23. Expected RAROC — Using 99.97% Co-CTE Allocation

	Expected Economic Profit	Allocated Allocated Capital	Expected RAROC
Line B	521,600	4,225,340	12.3%

You'd like to have an expected RAROC for this line of at least 25%. What should the new premium be, assuming that adjustments to the premium do not impact the allocated capital?

Solution. There are two ways that this can be done. The first is to just write the formula for RAROC including a term for the additional risk margin, then solve for this unknown.

The calculations would be as follows:

$$\begin{aligned} P &= \text{Original Premium} \\ E &= \text{Expenses} \\ i &= \text{Expected Investment Income} \\ PV(L) &= \text{PV of Expected Claims} \end{aligned}$$

$$\begin{aligned} \text{RAROC} &= \frac{[P + \pi - E](1 + i) - PV(L)}{\text{Allocated Risk Capital}} \\ &= \frac{[6{,}400{,}000 + \pi - 320{,}000](1 + 5\%) - 5{,}862{,}400}{4{,}225{,}340} \\ &= 25\% \\ \pi &= 509{,}271 \end{aligned}$$

So the new premium will be $P + \pi = 6{,}909{,}271$.

Another way to do the same calculation is to note that if the \$4,255,340 is treated as the required capital to write the business and the 25% RAROC target is the per unit cost of capital, then the total dollar cost of the capital is:

$$\text{Dollar Cost of Capital} = 25\%(\$4{,}255{,}340) = \$1{,}056{,}335$$

This is the amount of expected economic profit that would have to be incorporated into the premium. Since the original premium already accounted for \$521,600 of this expected profit, only \$534,735 of additional expected profit would be required at the end of the year. Discounting that amount to reflect the fact that it is collected up-front, \$509,271 of additional risk margin would have to be incorporated into the premium to meet the RAROC target rate.

TABLE 24. Calculation of Additional Risk Margin Required for 25% RAROC

		Amount	Calculations
(1)	Allocated Risk Capital	4,225,340	Based on Co-CTE Allocation
(2)	Target RAROC	25.0%	Assumed
(3)	Required Economic Profit	1,056,335	(3) = (1) * (2)
(4)	Current Economic Profit	521,600	Based on Assumptions
(5)	Shortfall	534,735	(5) = (3) - (4)
(6)	Expected Investment Income	5.00%	Based on Assumptions
(7)	Additional Risk Margin Required	509,271	(7) = (5)/[1 + (6)]

Question 9. Suppose you are given simulated underwriting results for three lines of business and the combined total, but each of these four sets of values have been sorted independently

already. How would you calculate the allocated capital to each line of business using the Co-CTE allocation method?

Solution. The short answer is, you can't do it! To calculate the co-measures you need to sort on the total but know which result from the three lines corresponds to the worst total outcomes. If each of the data sets have already been independently sorted, you cannot apply the Co-CTE method directly.

Note that I put this question in here to address what appears to have been a defective question on the 2014 exam (Question 12c).

Question 10. ABC Reinsurance Company is considering whether to write one of the following portfolios of property exposures:

- Portfolio A contains 100 locations with a combined total insured value of 45
- Portfolio B contains 250 locations with a combined total insured value of 160

Below are some summary statistics for the original ABC portfolio as well as the effect of adding either Portfolio A or Portfolio B to the existing portfolio:

	ABC Co.	Portfolio A	Portfolio B	ABC + Portfolio A	ABC + Portfolio B
Variance of Claim Costs	2,538.69	15.92	383.13	2,581.54	4,836.37
Coeff. Of Variation of Claim Costs	4.14	8.62	4.4	4.03	4.19
Expected Profit	3.20			3.38	4.28
Total Risk-Adjusted Capital	20.00			22.00	25.00

The company can write only one of these portfolios, A or B, not both. They require a 15% return on risk-adjusted capital (RAROC) after adding the new risk. Fully discuss the elements ABC Reinsurance Company should consider when deciding which portfolio to write.

Note that this is Question 22b from the 2014 exam, but with some additional information from the question that was needed for Part (a) removed.

Solution. First, notice that whether they write A or B, either one will satisfy the 15% RAROC requirement. After adding Portfolio A to the existing portfolio, the RAROC is $3.38/22 = 15.4\%$; after adding Portfolio B to the existing portfolio the RAROC is $4.28/25 = 17.1\%$. However, you need to be careful about two important points discussed by Feldblum in his IRR paper:

- When comparing two mutually exclusive investment opportunities, you cannot just compare the RAROC ratios and conclude that the higher one is "better". Often the size of the investments will differ and earning a high return on a small investment may not be better than earning a lower return on a smaller investment.

- When comparing two possible portfolios to add to your existing portfolio, you need to ensure that the marginal impact on the profit or RAROC is attractive. If the combined portfolio still meets the 15% hurdle rate but results in a reduction of the RAROC compared to the original portfolio, you may choose to not make the investment.

Because of these considerations, you should look at the economic value added (EVA). You can first compare the EVA for the two options: for Portfolio A, the EVA is $3.38 - 15\% * 22 = .08$ and for Portfolio B the EVA is $4.28 - 15\% * 25 = .53$. Given these two choices, adding Portfolio B is better. And when compared to the EVA of simply sticking with the existing portfolio ($3.20 - 15\% * 20 = 0.20$), this is a marginal improvement and so Portfolio B should be added.

You could argue, as some did on the exam (see the Examiners' Report), that the riskiness of the two portfolios under consideration (e.g. as measured by their stand-alone variance), the variance of the two possible combined portfolios, and the covariance of the two portfolios under consideration with the existing portfolio should also be factored into the decision. For instance, it appears from the information provided that Portfolio A has very low correlation with the existing portfolio, since the combined variance is only slightly higher than the original variance. In contrast, Portfolio B appears to be highly correlated with the existing portfolio as the combined variance increases dramatically.

Those are indeed important considerations, but I would argue that they are already captured in the "Total Risk-Adjusted Capital" measure provided in the question. By measuring the EVA as I did above, I am trying to balance the additional expected profit from taking on more risk against the capital required to support this additional risk, and to find the option that gives me the best reward per unit of risk. Simply saying that Portfolio B is riskier is not sufficient to evaluate this trade-off.

Selected Old Exam Questions for Part 5

The following questions relevant for this section appeared on the Old CAS Exam 8 from 2000 to 2010 and on the CAS Exam 9 since 2011.

Stulz	**Culp**	**Butsic**	**Cummins Alloc.**	**Goldfarb**
2003 Q47	2003 Q45	2000 Q44A	2003 Q18	2007 Q35
2005 Q39	2005 Q38	2001 Q46	2005 Q40	2008 Q35
2006 Q37	2006 Q36	2002 Q40	2005 Q41	2008 Q36
2007 Q34	2007 Q33	2002 Q41	2006 Q38	2008 Q37
2010 Q26	2012 Q9	2004 Q39	2008 Q38	2009 Q31
2014 Q17		2006 Q39	2010 Q28	2010 Q27
		2007 Q36	2011 Q11	2011 Q13
		2008 Q39		2011 Q14
		2008 Q40		2012 Q10
		2009 Q33		2013 Q15
		2010 Q30		2013 Q16
		2013 Q14		2014 Q12
		2015 Q16		2014 Q14
				2014 Q22b
				2015 Q15
				2015 Q17
				2015 Q18

For some of these questions I have provided the text of the question and an explanatory solution. These were selected either because they are representative of the questions you are likely to be asked on future exams or because they contain an element that is particularly worthwhile to point out. For the other questions, the CAS solutions should be sufficient to confirm whether your answer is correct.

Important Note: The solutions shown here are intentionally detailed. They contain thorough explanations of the concepts and formulas used to reinforce the main points from the readings and provide an additional teaching opportunity. **Actual exam responses should be much more concise than what is shown here, along the lines of what you will see in the solutions that the CAS releases.**

2007 Exam Question 36

You are given the following:

- Assets = \$1,000
- Liabilities = \$1,196
- Risk free rate (continuously compounded) = 6%
- Time to payment = 1 year
- Risk parameter (continuously compounded) = 10%

Calculate the expected policyholder deficit.

Notice that in the Butsic paper he was always careful to identify whether the assets were stochastic and the liabilities were fixed or the assets were fixed and the liabilities stochastic. Importantly, his formulas differed depending on which of these cases he was dealing with.

In the Cummins paper though, he allowed both the assets and the liabilities to be stochastic and focused on the asset to liability ratio rather than either the assets or the liabilities. Then he was able to calculate the *EPD Ratio* using the standard Black-Scholes put formula.

So in Cummins, the EPD Ratio is a standard put option on the ratio of the assets to the liabilities with a strike price equal to 1.0. If the ratio of assets to liabilities is less than 1.0 on the maturity date, then the firm is insolvent and the deficit (as a percent of the liabilities) is the amount by which the ratio is below 1.0. The volatility parameter in the Cummins examples is therefore the volatility of the asset-to-liability ratio. It reflects not just the separate asset and liability standard deviations but also their correlations. (Notice that the question used the term *risk parameter* which comes directly from the Cummins paper.)

One important catch though is that Cummins also assumed that $r = 0$ (as did Butsic) and treated the current liability value as the forward value of the liability. For simplicity here, I will assume there is nothing complicated to worry about and just go ahead and use the 6% risk free rate that was specified (see below for the details that are being omitted for the moment).

The put can be valued as $p = 0.1107$ using the standard Black-Scholes formulas as follows:

Assets	1,000
Liabilities	1,196
A/L Ratio	0.8361
Strike Price	1.00
Risk free	6.00%
Volatility	10.00%
Expiration	1.00
d_1	-1.1398
d_2	-1.2398
$N(d_1)$	0.1272
$N(d_2)$	0.1075
Call	0.0051
Put	0.1107
EPD Ratio	0.1107
EPD	132.42

Notice that the question asked for the EPD and not the EPD Ratio. This can be obtained by multiplying the EPD Ratio of 0.1107 by the liability amount to get EPD = 132.42.

Subtle Complication

First, the sample solution suggested that the strike price is the present value of the liabilities. In fact, the strike price is the *forward* value of the liabilities — that is the amount that has to be paid at maturity, not the present value.

More importantly, when the assets and liabilities are both stochastic, what we are really doing is valuing the option to exchange the assets for the liabilities or put the assets to the policyholders in exchange for extinguishing the liability. The call version of this, where someone has the option to get the value of the assets if they wind up being worth more than the liabilities can be valued using a formula shown in a chapter of Hull that is not on the exam:

$$\text{Option to Get Assets in Exchange for Liabilities} = A_0 N(d_1) - L_0 N(d_2)$$

where,

$$d_1 = \frac{\ln(A_0/L_0) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \qquad d_2 = d_1 - \sigma\sqrt{T}$$

Here, it is critical to recognize that A_0 is the current value of the assets and L_0 is the current value of the liabilities. Since Cummins used L to signify the *forward value* of the assets $L_0 = Le^{-rT}$. If we want to write the option value in terms of L we have to write this as:

$$\text{Option to Get Assets in Exchange for Liabilities} = A_0 N(d_1) - Le^{-rT} N(d_2)$$

where,

$$d_1 = \frac{\ln\left(\frac{A_0}{Le^{-rT}}\right) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}$$
$$= \frac{\ln\left(\frac{A_0}{L}\right) + \left(r + \frac{1}{2}\sigma^2\right) T}{\sigma\sqrt{T}}$$

Now we have to make two additional changes. First, we can write this as the ratio of the option value to the liability value:

$$\text{EPD Ratio} = \frac{A_0}{L}N(d_1) - \frac{Le^{-rT}}{L}N(d_2)$$
$$= \frac{A_0}{L}N(d_1) - (1.0)e^{-rT}N(d_2)$$

Now we can see that this looks like a call option on the ratio A_0/L with a strike price of 1.0.

Finally, note that this is the value of what is essentially a call where the holder receives the assets if they exceed the liabilities. We want the value of the put where we can give someone the assets if they are worth less than the liabilities. We can just use put-call parity for that, as was done in the numerical calculation above.

$$\text{EPD Ratio} = [\frac{A_0}{L}N(d_1) - (1.0)e^{-rT}N(d_2)] + (1.0)e^{-rT} - \frac{A_0}{L}$$
$$= \frac{A_0}{L}[N(d_1) - 1] + (1.0)e^{-rT}[1 - N(d_2)]$$
$$= \frac{A_0}{L}[-N(-d_1)] + (1.0)e^{-rT}[N(-d_2)]$$
$$= (1.0)e^{-rT}N(-d_2) - \frac{A_0}{L}N(-d_1)$$

That is the formula for a put on the asset-to-liability ratio $\frac{A_0}{L}$ with a strike price of 1.0.

So now you can see that the only reason the calculation we did above was correct was that the liability value given ($1,196) was assumed to be the forward liability value. That's why it made sense to include the risk free rate term in the d_1 term and to discount the strike price. If the liability value were actually the current (present) value of the liability, then we would have to ignore the risk free rate term in both cases.

2010 Exam Question 27

ABC Insurance Company held a portfolio of workers' compensation (WC) losses with reserves equal to \$35,000 as of December 31, 2008. Management uses a Co-Measure Conditional Tail Expectation (Co-CTE) at the 95% threshold to allocate surplus to risk sources.

XYZ Insurance Company is in runoff and transferred a portfolio of medical professional liability (MPL) losses with reserves equal to \$25,000 to ABC Insurance Company. XYZ also transferred to ABC assets equal to the MPL reserves plus a risk reserve equal to the stand-alone CTE at the 95% threshold for the MPL reserves.

Below is ABC Insurance Company's balance sheet before and after the loss portfolio transfer, which took place on January 1, 2009:

	Assets	Liabilities	Actual Capital
December 31, 2007	155,000	35,000	120,000
December 31, 2008	155,000	35,000	120,000
December 31, 2009	?	60,000	?

Assets are invested at the beginning of each year and investment income is paid at the end of each year. 100% of investment income is paid out as dividends. Return on invested assets was 2.5% in 2008 and 2.0% in 2009.

An outside consultant modeled the aggregate loss from three risk sources (asset risk, WC underwriting risk and MPL underwriting risk) before and after the portfolio transfer. Results before the transfer:

	Asset	WC	MPL	Total
ABC Insurance Company Stand-alone CTE at 95%	40,000	68,000	n/a	n/a
XYZ Insurance Company Stand-alone CTE at 95%	32,000	n/a	75,000	n/a
ABC Insurance Company Co-CTE at 95%	35,000	75,000	n/a	110,000
XYZ Insurance Company Co-CTE at 95%	36,000	n/a	66,000	102,000

100 scenarios were simulated to model the aggregate results after the transfer. The 10 worst results of these 100 scenarios folllow:

Scenario	Asset	WC	MPL	Total
33	45,500	156,000	104,000	305,500
89	39,000	117,000	130,000	286,000
69	32,500	104,000	130,000	266,500
77	52,000	65,000	143,000	260,000
67	19,500	182,000	52,000	253,500
30	26,000	13,000	182,000	221,000
97	19,500	65,000	130,000	214,500
88	13,000	156,000	26,000	195,000
70	-1,300	104,000	91,000	193,700
75	-6,500	65,000	117,000	175,500

a. Calculate ABC Insurance Company's assets and actual capital (surplus) at December 31, 2009.

Here, we just need to know the assets that were transferred from XYZ to ABC. We are told that there was a risk reserve equal to the 95% CTE. The CAS sample solution (presumably the full credit solution) took this literally. However, since we are preparing for an exam covering risk load methodology, does it make sense that a risk load would be calculated as the full amount of the tail of the distribution (approximately the 97.5th percentile)? That means that they paid the mean loss *plus* the 97.5th percentile amount. That could be more than the maximum possible loss!

Since I find that to be silly[1], I am going to work through the rest of my answer assuming that what they meant to say is that they transferred assets equal to the 95% CTE, some of that consisting of the 25,000 in reserves and the rest reflecting a risk margin. All of my answers will therefore differ from the CAS answer, but the differences should be clear. Refer to their answer if you care.

In this case, the 95% CTE for XYZ's MPL book, on a stand-alone basis was 75,000. This is the total amount of assets transferred. Their ending assets would have been $155,000 + 75,000 = 230,000$ and their liabilities are shown to be $60,000$ so that gives $230,000 - 60,000 = 170,000$ in surplus.

b. Calculate the actual capital allocated to each risk source after the portfolio transfer for ABC Insurance Company using the 95% Co-CTE allocation method.

First we need to calculate the Co-CTEs. We do this by sorting the results of the simulation based on the Total column, in descending order, and using the worst 1-95% = 5% of the scenarios. Since 100 simulation scenarios were run, we use the worst 5. Then using those 5 scenarios we simply calculate the average in each column. This amount will sum to the average in the total column and can be used as a proportional allocation of the total surplus.

[1]And yes, I do see the irony that had I worked this out on the actual exam I would have gotten every part wrong. This is a question based on a paper that I wrote and I wouldn't have been given full credit by the CAS. You have to love that!

If their ending surplus is 170,000 and they allocated this to the three risk categories in proportion to their 95% Co-CTEs, then we will have the following:

Scenario	Asset	WC	MPL	Total
33	45,500	156,000	104,000	305,500
89	39,000	117,000	130,000	286,000
69	32,500	104,000	130,000	266,500
77	52,000	65,000	143,000	260,000
67	19,500	182,000	52,000	253,500
Average	37,700	124,800	111,800	274,300
% of Total	13.74%	45.50%	40.76%	100.00%
Allocated Capital	23,365	77,346	69,289	170,000

Notice that the question referenced *underwriting risk* while the information in the question implies that these are only *reserve risks* that we are dealing with. No information was given about ABC's on-going underwriting of new business. While we might assume a steady-state and therefore treat the CTE information for ABC as being the underwriting risk, it wouldn't make sense for the CTE calculations to have ignored the reserve risk. And for XYZ, they are in run-off and so their risk must truly have been reserve risk and not underwriting risk as those terms were defined in the paper.

c. Calculate the investment income earned in 2008 and 2009.

We are told that the investment yields were 2.5% and 2.0% and so the investment income, assuming no taxes, would be:

	2008	2009
Assets	155,000	230,000
Investment Income Rate	2.5%	2.0%
Investment Income	3,875	4,600

d. Calculate the risk-adjusted return on capital (RAROC) for ABC Insurance Company's asset risk as of December 31, 2008 and as of December 31, 2009.

To calculate the RAROC we simply have to divide the income by the allocated surplus. We already calculated the allocated surplus for 2009 in Part (b).

What about for 2008? Here, there is no reason to assume that the allocation has to be done the same way, but it is logical to be consistent and assume that the 2008 allocation of capital would also be done using the 95% CO-CTE.

This calculation is:

	Assets	WC	Total
95% Co-CTE	35,000	75,000	110,000
% of Total	31.82%	68.18%	100.00%
Allocated Capital	38,181.82	81,818.18	120,000.00

Since we want the RAROC attributable to the assets, we can use just the investment income in the numerator and the allocated capital to the assets in the denominator, which gives us:

	2008	2009
Investment Income	3,875	4,600
Allocated Capital	38,182	23,365
RAROC	10.15%	19.69%

2010 Exam Question 28

An insurance company writes three lines of business with the following characteristics:

- The partial derivative of a put option value, p, with respect to the capital to liability ratio, s, is
$$\frac{\partial p}{\partial s} = -.025$$
- The partial derivative of a put option value, p, with respect to the standard deviation, σ, is
$$\frac{\partial p}{\partial \sigma} = .085$$
Note: They meant the σ parameter of the loss distribution, not the standard deviation.
- Assume assets and liabilities are independent.
- The expected loss and income for each line are below:

Line	Expected Loss	Expected Income	Covariance with Total Losses
1	1,500	25	$\sigma_{1L} = .0150$
2	2,250	65	$\sigma_{2L} = .0225$
3	4,000	100	$\sigma_{3L} = .0275$
Total	7,750	190	$\sigma_L^2 = .0236$

- The following capital levels are calculated using an Expected Policyholder Deficit target ratio of 5% for the following lines and line combinations:

Line(s)	1	2	3	1 & 2	1 & 3	2 & 3	1, 2 & 3
Capital	750	950	1,375	1,224	1,615	1,860	1,999

a. Calculate the allocated capital for each of lines 1, 2, and 3 using the Merton-Perold marginal capital allocation method.

Under the M-P allocation method, using the capital levels that produce a 5% EPD ratio as the basis, we allocate capital to each line based on its marginal contribution to the total capital needed for all lines combined.

For instance, without line 1 the firm would need \$1,860 of capital (the combination of lines 2 and 3) and with line 1 they need \$1,999. This means the marginal capital is \$139. The rest are shown below:

Line	Marginal Capital
1	139
2	384
3	775

a. Calculate the allocated capital for each of lines 1, 2, and 3 using the Myers-Read marginal capital allocation method.

First, let's just write out the full formula as shown in the Cummins reading for the Myers-Read surplus to liability ratios, s_i:

$$s_i = s - \left(\frac{\partial p}{\partial s}\right)^{-1} \left(\frac{\partial p}{\partial \sigma}\right) [(\sigma_{iL} - \sigma_L^2) - (\sigma_{iV} - \sigma_{LV})] \frac{1}{\sigma}$$

In that formula, the σ terms with the V subscripts reflect the covariance of the line i losses with the assets and the covariance of the total losses with the assets, respectively. Since the assets and liabilities were assumed to have no correlation, these terms drop out.

The σ parameter with no subscripts reflects the overall volatility of the assets and liabilities. In the Cummins reading he assumed assets had no volatility, so $\sigma = \sigma_L$. Here, all they said was that the assets and liabilities were uncorrelated, not that the assets were fixed. We have no other way to proceed, so let's assume $\sigma = \sigma_L = \sqrt{.0236} = 0.15362$.

Finally, the term σ_{iL} reflects the covariance of line i with the total losses for all lines, which are the values given in the table above.

We can now just proceed to filling the right values into the formula, with one calculation for each line. We just need the firmwide capital to liability ratio,

$$s = \frac{1,999}{7,750} = 0.25794$$

σ_L	0.15362
$\partial p / \partial s$	-0.02500
$\partial p / \partial \sigma$	0.08500
s	0.25794

Line	Expected Loss	Covar	s_i	Capital
1	1,500	0.0150	0.0676	101.40
2	2,250	0.0225	0.2336	525.58
3	4,000	0.0275	0.3443	1,377.00
Total	7,750	0.0236	0.2579	1,999.00

The Myers-Read formula gives the capital to liability ratios by line. We get the allocated capital by multiplying by the expected losses by line.

c. The company has a hurdle rate of 10%. Fully explain how the Merton-Perold and Myers-Read allocation methods could lead to different business conclusions.

The M-P allocation does not add to the total capital for the firm. In this case, only 1,298 of the total 1,999 capital required was allocated. This will make some or all of the lines' reported return on allocated capital larger than if all the capital was allocated and make it hard to interpret accurately.

This isn't an issue with the M-R allocation.

d. Using the allocated capital calculated for the two methods above, show which lines would be accepted or rejected under both allocation methods.

If the 10% hurdle rate were applied in both cases, both Lines 1 and 2 would be accepted by both methods. For Line 3, the M-P method would accept it but the M-R would reject it.

	Allocated Capital			Return on Capital	
Line	M-P	M-R	Income	M-P	M-R
1	139	101	25	18%	25%
2	384	526	65	17%	12%
3	775	1,377	100	13%	7%

2011 Exam Question 13

An insurer uses a risk-adjusted return on capital (RAROC) framework to evaluate its pricing adequacy for a given line of business. Given the following information about the line of business:

- Discounted expected loss and LAE = 60,000,000
- Undiscounted expected loss and LAE = 64,000,000
- Expense ratio = 30%
- Total investment return = 8%
- Target risk-adjusted return on capital = 15%
- Allocated capital to the line = 20,000,000
- All premium and expenses are collected and paid at the beginning of the year
- Losses are paid at the end of the year

a. Determine the combined ratio the insurer would need to achieve for this line in order to meet its RAROC target.

In the Goldfarb paper, the formula used was:

$$\text{RAROC} = \frac{[P - E](1 + i) - PV(L)}{\text{Allocated Risk Capital}}$$

This should be a simple use of this formula, except that the question made it unclear how to interpret the information given.

First, in the paper loss and LAE was always discounted *to the end of the first year* so that the term $PV(L)$ refers to the value of the losses as if they are all paid at the end of the first year. In the question, they said that all losses are paid in the first year, so technically $PV(L) = 64,000,000$. This is confusing because this amount is described as the undiscounted amount, but it appears as though the discounted amount of $60,000,000$ refers to the present value as of today because the losses are said to all be paid at the end of the year. This was unclear enough that the CAS appears to have accepted the use of either the undiscounted or the discounted loss amount.

Second, in the numerical examples in the reading it was assumed explicitly that the definition of income excluded the investment income that could be earned on the allocated capital and that the target risk-adjusted return on capital also excluded investment income (it was an excess return over the investment return). It's not clear from the question whether the 15% rate given was intended to be inclusive or exclusive of the 8% investment return. I will assume it was exclusive of the 8% rate and do the calculations just as in the paper.

Using the formula shown above, we simply solve for P:

$$15\% = \frac{[P(1-.3)](1.08) - 64,000,000}{20,000,000} \Rightarrow P = 88,624,339$$

Using this premium, the combined ratio would be 102.2%.

b. The underwriting results for this line are highly correlated with other business units within the insurer. Capital was allocated to the business units using a proportional allocation method. Fully explain whether the target combined ratio from Part (a) above would likely be higher or lower if capital were allocated based on the Merton-Perold methodology.

The M-P method takes into account the incremental impact of each business unit, which will tend to result in higher allocations of capital, and thus lower target combined ratios (i.e. higher required underwriting profits) for business units that are more highly correlated with the other businesses (collectively). Without knowing the stand-alone variability of each line, the size of the lines and the degree of correlation among the other lines of business it is actually difficult to say for sure whether the M-P method will result in higher or lower allocated capital than the proportional allocation method for any given line.

In addition, the M-P method often results in allocation of less than the company's total capital and there are different thoughts as to what should be done with the residual capital. If it isn't allocated, then some lines of business will necessarily get less capital allocated to them than under the proportional allocation method, thereby increasing the target combined ratio.

In the end, I think it's hard to say what will happen, which is why the CAS seems to have accepted "higher combined ratio" and "lower combined ratio" for full credit.

2011 Exam Question 14

An insurer has set a target return on capital of 12%. A prospective risk requires multiple years of capital investment. The capital allocated to the risk runs off at the end of each year according to the following schedule:

Year	Capital Released
1	30%
2	40%
3	20%
4	10%

The insurer determines it needs to allocated \$500,000 of initial capital to support the risk.

If the risk-free rate of return is 7%, calculate the required economic profit that must be earned from the risk in order to meet the insurer's required return on capital.

When capital is held for multiple periods, you simply need to account for the fact that investors will demand to earn some target return on the capital held in each period. The table below applies the 12% cost of risk capital each period to the capital held at the start of the period and discounts all amounts back **to time zero** at the **risk-free** rate. The total is the present value economic profit that must be earned:

Year	Beginning Capital	Cost of Risk Capital	PV Cost of Risk Capital	Capital Released	Ending Capital
1	500,000	60,000	56,075	150,000	350,000
2	350,000	42,000	36,684	200,000	150,000
3	150,000	18,000	14,693	100,000	50,000
4	50,000	6,000	4,577	50,000	0
Total			**112,030**		

Notice two important points.

First, the question originally referred to the 7% rate as the "interest income on the risk capital held each year". The reason for discounting the values above, at least the reason given in the paper, isn't to account for investment income on the capital but rather to account for the fact that you get to collect the risk margin *today*. This means that the amount you need to include in the premium is reduced by the risk-free investment income rate. To make this less confusing, I edited the wording of the question a bit.

Further, the question technically asked about the economic profit you need and not the premium you need to collect. In the paper economic profit was defined as of the end of the first year, so technically the discounting should have only been done to the end of the first period rather than to today.

2012 Exam Question 10

An insurer estimates the 99% Value at Risk for its three sources of risk as follows:

- UW Risk: 100m
- Reserve Risk: 80m
- Credit Risk: 70m

For combinations of these sources, the following estimates of the 99% Value at Risk are also given:

- UW and Reserve Risk: 120m
- UW and Credit Risk: 130m
- Reserve and Credit Risk: 140m
- UW, Reserve and Credit Risk: 180m

The insurer uses the 99% Value at Risk to estimate its risk capital.

a. Calculate the portion of total risk capital allocated to UW risk using the proportional method.

Using each of the stand-alone 99% VaR amounts, UW risk comprises 100/250 = 40% of the total. Allocating this portion to UW risk would give UW risk capital equal to 72m.

Calculate the portion of total risk capital allocated to reserve risk using the incremental method.

One version of the incremental method is the Merton-Perold method, which would determine the total risk capital with all three risk sources and subtract the risk capital without the reserve risk. The difference, or $180m - 130m = 50m$ would be the allocation to reserve risk.

Note that the sample solution given in the examiners report then rescaled the incremental amounts so that they summed to the total amount of 180m. This is not how Merton-Perold used this method and instead they left a portion of the capital unallocated.

Decide whether the insurer should allocate its total estimated risk capital or its actual book value to the risk sources. Give two reasons for your decision

There are many arguments on both sides (see the CAS Examiner's Report), but in the Goldfarb reading the argument is made to allocate the book capital because it is the actual amount of capital on which the investors want to earn returns. In addition, it is easier to determine a fair target return on capital when book capital is used because it is not subject to leverage effects which inflate the return on risk capital that arise because of the arbitrary way in which we can define risk capital.

2012 Exam Question 11

An insurance company writes two lines of business:

	Premium	Loss & LAE Reserves	Beta of Liabilities
Line A	500	600	0.25
Line B	500	200	0.35

- The company has \$1 billion of surplus and no liabilities other than loss and LAE reserves
- The company's assets have a beta of 0.50
- The risk-free rate of return is 4%
- The equity market rate of return is 10%

a. Calculate the expected rate of return on the firm's equity.

This is covered in the Cummins reading, but could be figured out without really having read it there. The firm's total profits are made up of investment income on its assets and underwriting income on its two lines of business.

$$I = r_A A + r_1 P_1 + r_2 P_2$$

Note that the underwriting return is measured as an underwriting profit relative to premium. If we express that total income relative to equity, we have the following expression:

$$I/E = r_E = r_a(A/E) + r_1 P_1/E + r_2 P_2/E$$

Substituting $A = E + R_1 + R_2$ where R_i is the reserve balance, we can rewrite this in terms of leverage ratios (k_i = reserve to surplus or liability leverage and s_i = premium to surplus or premium leverage):

$$r_E = r_A(1 + k_1 + k_2) + r_1 s_1 + r_2 s_2$$

We can then write this in terms of betas rather than returns:

$$\beta_E = \beta_A(1 + k_1 + k_2) + \beta_1 s_1 + \beta_2 s_2$$

From this, we can calculate the equity beta as:

$$\beta_E = .5(1 + .6 + .2) + .25(.5) + .35(.3) = 1.2$$

Plugging this into CAPM gives $r_E = .04 + 1.2(.10 - .04) = 11.2\%$.

b. Calculate the target combined ratios for the two lines of business so that the equity return is fair.

Note that the question originally just asked for the combined ratios, but I have reworded it to capture what I believe they meant.

To get to the accepted answer, I'll begin by just specifying the formula from the Cummins reading, which I will show below isn't correct. In the reading, the fair profit margin relative to premium (the underwriting return) was shown as:

$$r_U = -kr_f + \beta_U(r_m - r_f)$$

Here, Cummins defined k as the ratio of (steady state) reserves to equity. Using the values given in the problem, we would have the following target underwriting returns for each line:

$$r_1 = -600/1000(.04) + .25 * (.1 - .04) = -.9\%$$

$$r_2 = -200/1000(.04) + .35 * (.1 - .04) = 1.3\%$$

Converting these to combined ratios, we have 100.9% and 98.7% (the combined ratio is one minus the profit margin).

The above is what the graders expected. But there is a small problem with the formula in Cummins. It is based on the Fairley paper but uses a different definition for k. In the Fairley paper, k was the proportion of **premium** that can be invested and it was typically defined as the ratio of reserves to premium. Conceptually what Cummins did was right, but by defining $k = L/E$ his formula is inconsistent with the Fairley paper it is based on.

Here's what Fairley did using just a single line of business to keep it simple. He started with the same formulas we used in part (a) and showed the equity return can be written as:

$$r_E = r_A(1 + ks) + r_U s$$

where s is the ratio of premium to equity and k is the proportion of premium that represents policyholder funds that can be invested.

Then he wrote the above in terms of betas, just as before:

$$\beta_E = \beta_A(1 + ks) + \beta_U s$$

Now we can use CAPM to plug in for r_A in the first equation:

$$r_E = [r_f + \beta_A(r_m - r_f)](1 + ks) + r_U s$$

Similarly, we can plug the formula for β_E into CAPM to get another expression for r_E:

$$r_E = r_f + [\beta_A(1 + ks) + \beta_U s](r_m - r_f)$$

Now, just set these two equations for r_E equal to each other and solve for r_U:

$$r_U s = -r_f ks + \beta_U s(r_m - r_f)$$

$$r_U = -kr_f + \beta_U(r_m - r_f)$$

Notice that the key step at the end was to divide by s, which is why it actually is important that k be defined as a proportion of the premium. Or, to put it differently, $k = Reserve/Premium$, a reserve-to-premium ratio. In this question, the Examiner's Report used the ratio of reserves to equity, $k_1 = .6$ and $k_2 = .2$. In fact, it should have been $k_1 = 1.2$ and $k_2 = .4$ when both are defined relative to premium.

c. Describe two potential problems with using CAPM to set profitability targets for the lines of business.

Underwriting betas are very hard to calculate, especially for multi-line insurers. In addition, CAPM (and similar models) assume that only systematic risk is priced, whereas insurers may find it appropriate to price for skewness or other factors. And finally, recent evidence suggests that CAPM is an incomplete model even for the pricing of systematic risk (e.g. Fama-French model is now the standard asset pricing model).

2013 Exam Question 14

A company currently writes coverage for Line A and is considering expanding into Line B. For both lines, all losses will be paid at time 1. The table below shows the probability distribution of nominal loss amounts.

Probability	Loss - Line A	Loss - Lines A and B
0.4	0	0
0.3	6,000	11,000
0.1	8,000	12,000
0.1	0	12,000
0.1	14,000	23,000

You are also give the following information about the company:

- All assets are invested in risk-free securities earning an annual rate of 2.04%
- Both lines are priced at an expected loss ratio of 100%
- The company pays no expenses
- To support Line A on a stand alone basis, the company requires $5,800 of risk capital at time 0
- The company uses a fixed expected policyholder deficit ratio to determine its required risk capital

Calculate the amount of additional capital the company would require at time 0 to support its expansion into Line B.

This is a simple calculation of the capital needed for a target EPD ratio, but we have to first determine the target EPD ratio. We get that from the fact that we know the initial capital of $5,800 with only Line A is sufficient to meet the EPD ratio target.

Start by calculating the expected losses for Line A, which come to $4,000. This amount will be collected as premium and, along with the $5,800 in starting capital, will earn the risk-free rate of interest and be worth $10,000. Given the distribution of losses, this will result in a deficit in only one instance, when losses are $14,000. The resulting EPD is $10\% * \$4,000 = \$400$ and the EPD ratio is 10%.

Now we move to the case where we write Line A and Line B and solve for the capital needed at time 0 to achieve the same EPD ratio. Doing this by hand is tricky, but if you just assume for a moment that there will only be a deficit in the $23,000 loss scenario, we can easily solve a simple algebra problem.

The average loss amount is \$8,000 and if we want the EPD Ratio to be 10% (as in the Line A only case) then we need the expected deficit to be \$800. Since the probability of a \$23,000 loss is 10%, the deficit in that scenario needs to be \$8,000. The capital, C, is then the amount so that:

$$23,000 - (8,000 + C) * 1.0204 = 8,000 \Rightarrow C = 6,700$$

Before proceeding, we should check that there is indeed no deficit in any of the other loss scenarios. There isn't, so we know that our simplifying assumption was okay. If there was a deficit in the other scenarios, we would have had to redo the algebra.

This amount, \$6,700, is the total capital needed for both lines, so the marginal capital needed to add Line B is \$900, the amount that the total requirement exceeds the initial requirement.

2014 Exam Question 14

A company writes two lines of business, A and B. Both lines have the same loss ratio, the same expense ratio, the same premium and the same annual cost of capital rate. The only difference between the two lines is that their losses are paid out at different rates.

The loss payment patterns and other relevant assumptions are shown below:

EOY	A	B
1	60%	80%
2	30%	20%
3	10%	0%
Risk-free interest rate	5%	
Cost of Capital rate	15%	
Loss Ratio (Undiscounted)	70%	
Expense ratio	25%	
Premium	43.00	
Allocated Capital (Beg. Of Yr 1)	25.00	

Assume the premium is collected at the beginning of the year and all losses and expenses are paid at the end of the year. The allocated capital is released as claims are paid.

Determine which line of business the company should grow using the risk-adjusted return on capital (RAROC) framework.

Notice that because capital is held over multiple years and released only as losses are paid, the multi-year capital commitment makes it difficult to use an annual rate of return calculation. But it is easy to calculate the dollar amount of required profit each year (the allocated capital times the cost of capital rate) and compare the total profits expected to be earned to this required amount.

Begin by calculating the present value expected profit (discounted to time zero) by discounting the loss and expenses to time zero and subtracting these amounts from the premium:

TABLE 1. Expected Present Value Profit

		Discounted (to Today)	
	Undiscounted	A	B
Premium	43.00	43.00	43.00
Expenses	10.75	10.24	10.24
Losses	30.10	27.99	28.39
PV Profit		4.77	4.37

Next, we need to compare those profits to the sum of the discounted annual cost of capital (in dollars). We do this by multiplying the capital allocated at the beginning of the year by the annual cost of capital rate and then discounting these amounts to time zero for consistency with the expected profit:

TABLE 2. Required Present Value Profit

			Required Profit			
	Required Capital		Undiscounted		Discounted	
EOY	A	B	A	B	A	B
0	25.00	25.00				
1	10.00	5.00	3.75	3.75	3.57	3.57
2	2.50	0.00	1.50	0.75	1.36	0.68
3	0.00	0.00	0.38	0.00	0.32	0.00
Total					**5.26**	**4.25**

To determine which line to invest in, we notice that Line A has higher present value profits, but also much higher required profits because the capital is released more slowly. Its expected profits fall short of the required amount $(4.77 < 5.26)$. In contrast, Line B's expected profits exceed the required profits $(4.37 > 4.25)$. As a result, we should grow Line B (so long as this relationship continues to hold).

2015 Exam Question 18

An insurance company writes three lines: Auto Liability (AL), Auto Physical Damage (PD) and Workers' Compensation (WC).

- The total required capital is $\$5,000,000$
- Capital is allocated based on a Co-CTE risk measure
- The expense ratio is 27% for each line of business
- The interest rate for discounting is 2.5%
- The cost of capital is 15%
- Expense are paid at the beginning of the year and losses are paid at the end of each year

Line	Premium	Undiscounted Loss Ratio	Loss Reserve Duration	99.5% Co-CTE
AL	5,000,000	70.0%	2.5	2,275,000
PD	2,000,000	62.5%	0.8	1,625,000
WC	8,000,000	77.5%	3.5	2,600,000
Total	15,000,000	73.0%	2.8	6,500,000

Determine whether the Workers' Compensation line of business adds value to the company on a risk-adjusted basis.

To use RAROC, we first need to calculate the economic profit.

In the Goldfarb paper, this is done by investing the net premiums for one period and discounting all claim payments to the end of the period. In this question, they provide the nominal loss ratio and the loss reserve duration, which we can interpret as the weighted average time to payment and treat the claims as if they are paid in a single payment at that time. So, if we were discounting the claims to time zero, we would discount at the 2.5% rate specified for 3.5 years. However, we need to discount to the *end* of the period, so we discount for 2.5 years and get a discounted loss ratio of 72.86%.

We also need to invest the net premiums, but the question didn't specify an investment rate. It merely said to discount the claims at 2.5%. Using that same 2.5% rate, the economic profit would be:

$$8,000,000(1-27\%)(1.025)-8,000,000(72.86\%)=157,161$$

The required capital for the WC line uses the \$5,000,000 of total required capital and allocates in proportion to the Co-CTE, giving required capital of \$2,000,000 (the WC line Co-CTE is 40% of the total CTE).

Combining these, we get a RAROC of 7.86%. Since this is less than the cost of capital (15%), we would say that the WC line does not add value on a risk-adjusted basis.

Note that we could have reached the same conclusion using the Economic Value Added (EVA) approach. There, we would have subtracted the cost of capital, in dollars, from the economic profit to obtain $\text{EVA} = 157,161 - 15\%(2,000,000) = -142,838$. Since the EVA is negative, we conclude that the WC line does not add value on a risk-adjusted basis.

Two points are worth noting here.

First, the question didn't say whether the capital is released at the end of the year or if it is held until all of the claims are paid in 3.5 years (or something in between). In the reading there is a discussion of multi-period capital commitments and the fact that the cost of capital would have to be reflected for all years for which capital is held.

Second, the CAS sample solutions are different than mine! In the sample solution, the losses were discounted to time zero and the report listed bringing all values to the end of the period and including investment income in the profit calculation as a "common error". One could debate how to define economic profit and EVA, but if the solution is supposed to be based on the reading then the CAS sample solution is incorrect. The reading does it differently.

Appendix: Normal CDF Table

TABLE 1. Cumulative Normal Distribution (Positive x)

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 2. Cumulative Normal Distribution (Negative x)

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Part 6

Internal Rate of Return

Part 6 Table of Contents

Felblum: Internal Rate of Return Model

Introduction

This reading covers the same IRR model discussed in the Robbin reading and provides several additional points that are relevant for the exam. I will briefly summarize them here, focusing solely on the three sections of the reading that are included on the current syllabus (Sections 1, 3 and 6).

Historical Underwriting Profit Margin

Early rating bureau pricing procedures used a fixed profit margin of 2.5% for workers' compensation and 5% for other lines. This approach suffered from three notable problems:

- Didn't consider the time value of money.
- Didn't respond to changes in the competitive market environment to capture current return expectations of, for instance, investors.
- Used sales as a rate base, which doesn't take into account the equity provided from the investors (owners).

IRR Method and the Investor Perspective

Many historical discounted cash flows models used for pricing, even those that used sophisticated financial theory to establish the risk-adjusted discount rate, took the policyholder perspective and looked only at the cash flows between the insurer and the policyholder (he refers to this as the "product market" point of view). By ignoring the investors' perspective, these methods cannot ensure that sufficient capital can be attracted to the insurance market in order to supply the insurance product in the first place (he refers to this as the "financial market" point of view).

Feldblum argues that the investor perspective is important because of the inefficiency of federal income taxes on investment income. When an insurance company holds assets in marketable securities it pays taxes on its investment income. But when those after-tax returns are paid to shareholders in the form of dividends, the shareholders pay tax again. Had the shareholders held those funds and invested on their own, they would have only had to pay tax on the investment income once. Because of this double-taxation problem, the flow of equity capital between the investors and the company does impact the price of insurance — the more equity capital provided, the larger is this frictional cost to shareholders.

Review of IRR Decision Rule

Basic corporate finance texts often propose using the "IRR Rule" to evaluate investment projects. To do this, we determine the discount rate that would have to be applied to all future cash flows so that the present value of those cash flows was just equal to the amount of the initial required investment. This can be written as follows:

$$\text{Initial Investment} = \frac{\text{CF}_1}{1+\text{IRR}} + \frac{\text{CF}_2}{(1+\text{IRR})^2} + \cdots + \frac{\text{CF}_N}{(1+\text{IRR})^N}$$

Once the IRR is determined, the IRR rule says that projects where the IRR exceeds a target rate (driven by the project's cost of capital) should be undertaken. Projects where the IRR is below the target rate should be rejected.

Alternatively, you could also just subtract the initial investment from the present value of all future cash flows, call this total the *Net Present Value*, and solve for the IRR that sets the Net Present Value equal to zero:

$$0 = -\text{Initial Investment} + \frac{\text{CF}_1}{1+\text{IRR}} + \frac{\text{CF}_2}{(1+\text{IRR})^2} + \cdots + \frac{\text{CF}_N}{(1+\text{IRR})^N}$$

With this formulation, we could define a nearly identical "NPV Rule". The NPV Rule says to calculate the net present value of the cash flows using the target rate or cost of capital and then accept projects that have a positive NPV. There are some special cases where the IRR Rule and the NPV Rule might not lead to the same conclusion, but as you can see they are generally quite similar. Later we will return to some special cases where care must be taken to apply the IRR rule correctly.

Applying the IRR Model in Insurance

When applying this model to price insurance transactions we focus on the investors' equity flows, which typically consist of committed surplus up-front and then eventual return of income and surplus over time. Unlike cash flow models that focus solely on the insurance flows (premium, loss, expense), these cash flows are used only indirectly in the IRR model because they impose constraints on the amount of surplus that can be returned to investors and its timing.

But herein lies a key challenge. How do we determine how much surplus must be committed up-front and when it can be released? These are critical inputs to the method and so care must be taken to make reasonable assumptions.

At the company level, the initial surplus required is not always clear. Regulatory capital requirements only set *minimum* levels and actual company surplus at any point in time is highly volatile. As a result, it is common to rely on industrywide surplus levels as the basis (e.g. through industry premium to surplus ratios). Using industry levels of surplus implicitly assume that at the industry level the amount of capital invested is appropriate.

But this only accounts for the initial surplus. It is also critical to determine how this surplus can be released. Often in practice it is assumed that the surplus can be released in proportion to paid losses, earned premium or some combination of the two. For instance, in many of Robbin's examples in his paper he assumed surplus was committed in proportion to written premium initially and then released when all of the premium is earned (he called this a "block" assumption).

And of course, all of these choices regarding surplus directly affect how much is allocated to any one line of business at a given time, and can greatly influence the results. This is important because in the end, unlike in other industries where the "initial investment" is known, the committed surplus is a theoretical value.

Surplus Allocation

As noted above, using the IRR model to price particular lines of business requires an allocation of the firm's overall surplus to lines.

Surplus is often allocated in proportion to premiums or reserves, or more commonly a combination of the two, and the choice matters quite a bit for the amount allocated to any given line and its estimated IRR. As more surplus is allocated for longer periods of time, the IRR declines. So a more appropriate method might be to allocate the surplus in proportion to the relative riskiness of the lines of business and to release it as the risk dissipates.

Note: The Goldfarb and Cummins readings both discuss ways in which such risk-based allocation might be performed.

The importance of a more risk-based allocation and release of surplus is highlighted best when thinking about specific policy types. We often talk about underwriting profit margins by line of business, but within lines of business the policy forms can vary widely (e.g. claims made vs. occurrence coverage, first dollar coverage vs. high layer excess). In each of these cases, the policy form can significantly impact the amount and timing of the risk assumed.

Pitfalls in IRR Analysis

Feldblum's paper concludes with a summary of some common pitfalls associated with the use of the IRR rule.

IRR Rule vs. NPV Rule

As previously noted, the IRR rule and the NPV rule are, in many cases, identical decision rules. But this is not always the case, especially when there are constraints on total resources (i.e. when an optimization is needed), where two projects are mutually exclusive so that only one of the two can be undertaken or when the cash flows show multiple sign reversals.

Constraints

Consider the fact that the IRR calculation gives no sense of the scale, whereas the NPV is measured in dollars and so the scale is always obvious.

When choosing between one of two projects, would a small project with a 40% IRR and a $1,000 NPV be preferred over an alternative with a 10% IRR but a $1,000,000 NPV? Probably not.

Sign Reversals

When the cash flows used show multiple sign reversals, the IRR method can be difficult to use. Luckily, this rarely happens in insurance applications, unless overly simplified assumptions are used that cause this to occur. In that case, the problem is easily rectified.

Mutually Exclusive Contracts

When only one project can be taken on and we have to choose from among many options, it can be misleading to focus solely on the IRR. But this shouldn't be a major concern.

First, there is an easy solution to this issue, which is to evaluate one of the projects as a baseline and then view each of the other options as having incremental cash flows. If the IRR rule is applied to these incremental cash flows we can determine whether it makes sense to switch from the baseline project to this alternative.

Second, part of the problem has to do with the fact that the IRR analysis implicitly assumes that funds collected early can be reinvested at the IRR rather than at the cost of capital. But in our application we are usually solving for the underwriting profit provision that will equate the IRR with the cost of capital, so this isn't really an issue for us.

Presentation of Results

One concern with the IRR model is that it isn't always clear how an IRR lower than the company's cost of capital really impacts the company. In contrast, with the NPV calculation the NPV in this case will be a negative number — a clear indication that rates are inadequate.

In the IRR model for insurance, larger losses often result in large required surplus, which we recall is a theoretical amount and not an actual amount. If within our model we allocate more surplus, we will also generate more investment income, driving the IRR closer to the investment yield. So even when the IRR is inadequate (below the cost of capital), it may still be positive and may even look reasonable to regulators.

For this reason, even when IRR analysis is used internally it may be better to show NPV analysis in rate filings.

Practice Questions

Question 1. Suppose that you charge a premium of $12,000 for a policy with $12,000 in expected claims that will be paid in exactly four years and $2,000 of expenses that will be paid immediately.

Assume the premium is fully earned immediately so that there is no unearned premium reserve but loss reserves are booked immediately. Surplus must be maintained at a 2:1 ratio of undiscounted reserves to surplus.

All asset are invested in a zero coupon bond with a 10% yield. What is the IRR of the equity flows assuming that none of the increase in the value of the zero coupon bond is recognized as part of surplus until the bond matures?

Solution. We first need to determine the equity flows, which in this case is simple because we really just have a flow into the firm at the beginning and a flow out, depending on the actual losses, at the end.

At inception, note that if we had no surplus we would take in $\$12,000$ in assets from the premium, have to pay $\$2,000$ in expenses and have to book loss reserves equal to $\$12,000$. We also know that we have to maintain surplus of $\$6,000$ from the 2:1 reserve to surplus ratio. The only way to achieve this is for investors to contribute $\$8,000$ in capital. That's our first flow.

If we invest all of our investible assets of $\$18,000$ in the the zero coupon bond it grows to $\$18,000(1.1^4) = 26,354$ at the end of four years. Out of this we pay the claims and are left with $\$26,354 - \$12,000 = \$14,354$ in surplus that can be paid to investors.

And note that because of what I said to assume about the treatment of the bond value, surplus remains constant at the required level and no equity flows can occur until the maturity date of the bond after the liabilities are paid.

The IRR is then found as:

$$0 = -8,000 + \frac{14,354}{(1+IRR)^4} \Rightarrow IRR = 15.74\%$$

Question 2. Using the same facts as in the previous question, determine the underwriting profit margin that should be used to achieve a 25% IRR. Assume that the expenses remain at $\$2,000$ even if the premium amount changes.

Solution. We will start with the previous formula and rewrite it in terms of the key variables:

$$\begin{aligned} 0 &= -8,000 + \frac{14,354}{(1+IRR)^4} \\ &= -\left[-(P-L-E) + L/2\right] + \frac{\{P - E + [-(P-L-E) + L/2]\}\,(1.1^4) - L}{(1+IRR)^4} \end{aligned}$$

The first term, $-(P-L-E)$ reflects the shortfall at inception that is caused by the UW loss (in this case, we earned the full premium at inception, established loss reserves and incurred the expenses). The second term, $L/2$, is the required surplus, which is half of the loss reserve. The total of these two reflects the initial investment required and is multiplied by negative one because it reflects a cash outflow for the investors.

The numerator of the next term reflects the end of period value at the 10% annual investment return of the total assets invested (consisting of the premium, less expenses, plus the capital commitment calculated above), less the losses paid. This represents the ending surplus available to the investors.

Notice something important about the invested assets portion of the last term. I wrote it out fully so that you can see its components, but it actually simplifies to $L + L/2 = 1.5L$. Why is this? Because I defined the initial surplus to be enough to cover expenses, establish the loss reserve and maintain the appropriate amount of surplus relative to loss reserves. Since all of the expenses are paid up-front, the invested assets will only be the assets equal to the loss reserves and the required surplus.

Plugging in for $E = \$2,000$ and $L = \$12,000$ we can set $IRR = 25\%$ and solve for $P = 14,121$. Since P reflects the expected loss plus the underwriting margin, we can easily solve for:

$$U_{\$} = 14,121 - 2,000 - 12,000 = 121$$

As a percent of premium this is .86%.

Question 3. Suppose we modified Question 1 so that the invested assets earn 10% per annum in cash (or, equivalently, that we can recognize the increase in the value of the zero coupon bond each period). This would create intermediate cash flows that could be paid out to the equity holders. Determine the IRR in this case. Assume the initial premium remains at 12,000.

Solution. Just as before, we need to initially raise 8,000 from investors in order to achieve the required 6,000 in surplus at inception. We then have to maintain that surplus, but all of the income that comes from the invested assets (18,000 in total invested to earn 10% will generate 1,800 per period) is free to be distributed to investors.

The IRR in this case can be calculated as:

$$0 = -8,000 + \frac{1,800}{1+IRR} + \frac{1,800}{(1+IRR)^2} + \frac{1,800}{(1+IRR)^3} + \frac{18,000 + 1,800 - 12,000}{(1+IRR)^4}$$
$$\Rightarrow IRR = 17.68\%$$

Notice that changing the timing of the cash flows improved the IRR, but didn't materially complicate the calculations — other than requiring you to use your calculators to enter the cash flows and solve the polynomial for IRR.

Investment income is one of the two ways that numerical problems like this can create interim cash flows for the investors. The other is changes in the required surplus, which we'll address in the numerical problems for the Robbin IRR reading.

Question 4. Using the assumptions from Question 3 above, determine the premium required to achieve a 25% IRR. Assume that expense remain constant at \$2,000 even if the premium amount changes.

Solution. Just as in Question 2, we just have to write the NPV formula in terms of the unknown premium amount and now solve for premium rather than the IRR. Conceptually this is trivial, but doing this on the actual exam could be difficult, as you will now see.

First, as we did before, we write the NPV formula in terms of the unknown premium amount, P. But now we need to be careful not to make this overly complicated. As suggested above in the answer to Question 2, the premium doesn't really affect anything other than the first cash flow because of the way it was calculated. The invested assets, which impact the investment income, are really just equal to the loss reserves plus the required surplus. This allows us to write the NPV equation as follows:

$$
\begin{aligned}
0 &= -8{,}000 + \frac{1{,}800}{1+IRR} + \frac{1{,}800}{(1+IRR)^2} + \frac{1{,}800}{(1+IRR)^3} + \frac{18{,}000 + 1{,}800 - 12{,}000}{(1+IRR)^4} \\
&= -\left[-(P-L-E) + L/2\right] + \frac{1{,}800}{1.25} + \frac{1{,}800}{(1.25)^2} + \frac{1{,}800}{(1.25)^3} + \frac{18{,}000 + 1{,}800 - 12{,}000}{(1.25)^4} \\
P &= 1.5L + E - \left[\frac{1{,}800}{1.25} + \frac{1{,}800}{(1.25)^2} + \frac{1{,}800}{(1.25)^3} + \frac{18{,}000 + 1{,}800 - 12{,}000}{(1.25)^4}\right] \\
&= 13{,}291.52
\end{aligned}
$$

I mentioned above that these problems can be very difficult. If you aren't careful to make sure that the initial surplus plus the initial premium covers the expenses and the loss reserves, you could wind up in a situation where the investment income earned depends on the premium. This will leave the premium term in the polynomial and make the algebra harder to solve.

Sometimes though, you can't avoid this. Sometimes the wording of the problem is such that the later cash flows do depend on the premium. For instance, if unearned premium reserves are held that affect the surplus requirement or if the surplus is a function of the written premium. The next problem will explore this situation.

Question 5. Use the same assumptions as the previous problem, but assume that the *initial* surplus has to equal 67% of the written premium (a premium to surplus ratio of 1.5) and that beginning at the end of the first period the surplus again has to be maintained at 50% of the loss reserve. Assume the expenses are fixed at $\$2,000$ and do not vary with the premium and that the investment income can be withdrawn annually as it is earned. Calculate the premium required to achieve a 25% IRR.

Solution. The fact that the surplus requirement is a function of the premium has two effects on the calculations. The obvious one is that it changes the initial surplus requirement. The second is that it affects the first period's investment income and how much surplus can be withdrawn at the end of the first period.

Now that it's getting a bit more complicated, I'll use a table to lay out the components of the cash flows. But I think it might be best to start with the ANSWER and then show how to solve for it on the actual exam when you don't have the benefit of just using Excel's Solver, like I did.

In the table here, I calculate each component of the flows to the equity investors as of the end of each period.

	Premium		Loss and LAE		Expenses		
Time	Earned	Paid	Incurred	Paid	Incurred	Paid	UW Income
0	13,665	13,665	12,000	0	2,000	2,000	-335
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	12,000	0	0	0

	Reserves				Total	Premium	Invested
Time	UEP	Loss and LAE	Expenses	Surplus	Liab & Surplus	Receivable	Assets
0	0	12,000	0	9,110	21,110	0	21,110
1	0	12,000	0	6,000	18,000	0	18,000
2	0	12,000	0	6,000	18,000	0	18,000
3	0	12,000	0	6,000	18,000	0	18,000
4	0	0	0	0	0	0	0

Time	UW Income	Investment Income	Change in Surplus	Equity Flow
0	-335	0	9,110	-9,445
1	0	2,111	-3,110	5,221
2	0	1,800	0	1,800
3	0	1,800	0	1,800
4	0	1,800	-6,000	7,800

If we now calculate the IRR of the equity flows in the last column of the bottom panel, we'll get an IRR equal to 25%.

Let's walk through the calculations, starting with the UW income. I show that with \$13,665 of initial premium, all of which is paid and earned up-front, we incur \$12,000 of losses and \$2,000 of expenses right away (note that in practice you wouldn't earn the premium and incur the losses immediately, but you would incur the expenses). This gives us an immediate UW loss of \$335 that eats into our surplus. If we have to maintain initial surplus of \$9,110 (\$13,665/1.5) at inception, we will also need to get additional initial capital from the investors to cover this immediate UW loss. This is shown in the first row of the bottom panel.

The second panel is used to monitor the invested assets. We start by making sure that we have accounted for all reserves that might exist (premium, loss and expense), which in this case is easy because it is just the \$12,000 of loss reserves from inception until the end of the fourth period. We also track the required surplus, which starts out at \$9,110 but then drops to \$6,000 for the rest of the time. The total of all reserves plus surplus, less any premiums receivable, is the invested asset balance.

The investment income in this case can be assumed to be 10% on the initial invested asset balance (in other cases we might assume losses are paid evenly throughout the period and that we earn income on the average balance).

The final panel shows the flows to investors. They get to take out, each period, the UW income (or pay in the UW loss), the investment income and the decrease in the required surplus.

Okay, that explains the answer. But how would I actually calculate the $13,665 initial premium by hand? I would fill in the same table but use variables where needed:

	Premium		Loss and LAE		Expenses		
Time	Earned	Paid	Incurred	Paid	Incurred	Paid	UW Income
0	P	P	12,000	0	2,000	2,000	P-12,000-2,000
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	12,000	0	0	0

	Reserves				Total	Premium	Invested
Time	UEP	Loss and LAE	Expenses	Surplus	Liab & Surplus	Receivable	Assets
0	0	12,000	0	$P/1.5$	$12,000 + P/1.5$	0	$12,000 + P/1.5$
1	0	12,000	0	6,000	18,000	0	18,000
2	0	12,000	0	6,000	18,000	0	18,000
3	0	12,000	0	6,000	18,000	0	18,000
4	0	0	0	0	0	0	0

Time	UW Income	Investment Income	Change in Surplus	Equity Flow
0	$P - 12,000 - 2,000$	0	$P/1.5$	$P - 12,000 - 2,000 + 0 - P/1.5$
1	0	$[12,000 + P/1.5](10\%)$	$6,000 - P/1.5$	$0 + [12,000 + P/1.5](10\%) - [6,000 - P/1.5]$
2	0	1,800	0	1,800
3	0	1,800	0	1,800
4	0	1,800	-6,000	7,800

The final column is all that we need to solve for P:

$$
\begin{aligned}
0 &= P - 12,000 - 2,000 + 0 - \frac{P}{1.5} + \frac{0 + [12,000 + \frac{P}{1.5}](10\%) - [6,000 - \frac{P}{1.5}]}{1.25} \\
&\quad + \frac{1800}{1.25^2} + \frac{1800}{1.25^3} + \frac{7800}{1.25^4} \\
&= P[1 - \frac{1}{1.5} + \frac{10\%}{1.5 * 1.25} + \frac{1}{1.5 * 1.25}] - 14,000 + \frac{10\%(12,000)}{1.25} - \frac{6,000}{1.25} \\
&\quad + \frac{1800}{1.25^2} + \frac{1800}{1.25^3} + \frac{7800}{1.25^4} \\
&= P[.92] - 12,571.52 \\
P &= 13,665
\end{aligned}
$$

Question 6. What are three of the problems noted by Feldblum with respect to the historical 5% profit load that had been used by rating bureaus?

Solution. The problems he noted were:

- Didn't consider the time value of money.
- Didn't respond to changes in the competitive market environment to capture current return expectations of, for instance, investors.
- Used sales as a rate base, which doesn't take into account the equity provided from the investors (owners).

Question 7. Many insurance pricing models focus solely on the underwriting cash flows and ignore the equity commitments of the investors. Describe why this is a problem and identify the characteristic of insurance companies that makes the investors' perspective important.

Solution. Not ensuring that rates include an adequate provision for returns to equity investors, for instance by carefully tracking the flow of equity capital to and from investors, makes it impossible to ensure that enough capital can be attracted to the industry to satisfy society's demand for insurance.

And the reason the equity flows are important in the first place is that in the insurance industry the capital committed by investors is primarily invested in marketable securities. This subjects investors to double taxation of the investment income, which is something that could be easily avoided if investors simply invested their capital directly in the same marketable securities. This imposes an additional cost that must be considered in the rates.

Question 8. Unlike other industries, the committed capital (surplus) needed to write insurance policies is not easily quantified and is essentially a theoretical value, especially when a multi-line and/or multi-state insurer must allocate its overall surplus to lines or states. Describe how the amount of initial capital is typically quantified and the methods used to determine how this capital will be released over time.

Solution. The most common approach for setting initial surplus levels is to rely on industry premium to surplus or reserve to surplus ratios, under the assumption that the industry overall must be in equilibrium and operate with adequate capital overall.

Regulatory requirements, even risk-based capital requirements, are usually not appropriate because they reflect minimum capital levels not optimal or target capital levels. However, this could be rectified by using industry ratios of actual capital to risk-based capital, which would capture the same benefits as using industry premium to surplus or reserve to surplus ratios.

But even with initial surplus estimated, how it is assumed to be released over time is critically important to the IRR method. Premium to surplus ratios could be used, but this will generally result in surplus being released too quickly. Reserve to surplus ratios could be used, which would ensure that surplus is committed over the life of the expected loss payouts and would effectively penalize longer tailed lines. Or, as is common, a combination of these two approaches could be used.

Better still is to try to capture relative risk and release surplus as the risk dissipates.

Notice though that whatever methods are used will have a significant impact on the relative allocation of firmwide surplus across different lines of business, especially over the life of the loss payments as the premium earning patterns, reserve payout patterns and outstanding risk levels vary considerably across lines.

Question 9. Feldblum argues for a risk-based method for determining the release of capital. Give an example of two lines for which this method would produce wildly different results compared to the use of reserve to surplus ratios.

Solution. As one of many possible examples, consider a company writing workers' compensation and high layer excess casualty business with equal premium volumes.

While it may be true that, especially in later development years, that the workers' compensation line may have considerably more carried reserves at any point in time, it is likely that the outstanding risk in the excess casualty line is far greater. Much of the workers' compensation reserves may reflect readily estimated lifetime medical claims while the smaller excess casualty reserves may reflect low expected losses but with potential for multiple IBNR claims which could eventually prove to be full-limit claims.

IRRs calculated using reserve to surplus ratios for these two lines and IRRs using a risk-based surplus allocation are likely to produce considerably different results.

Question 10. Feldblum argues that even within a given line of business it may be appropriate to vary the amount of surplus assumed for different segments of the book. What specific issue does he raise?

Solution. He argues that the *policy form* can significantly influence the amount of risk in a given policy relative to another in the same line of business. For example, claims made vs. occurrence forms, retrospectively rated vs. fixed premium policies and first dollar coverage vs. high layer excess policies would have significant differences in surplus allocations if done on a risk basis.

Question 11. IRR decision rules are often criticized because they can, if used inappropriately, lead to incorrect decisions in special cases such as when choosing from among a collection of mutually exclusive projects (meaning they cannot all be done). Part of this problem relates to the fact that the IRR method implicitly assumes that interim cash outflows can be reinvested at the IRR. But Feldblum argues this isn't a problem in insurance applications of IRR. What is his point?

Solution. The implicit assumption that interim cash flows can be reinvested at the IRR creates problems when the IRR is greater than the cost of capital because each of these hypothetical reinvestments at the IRR are treated as adding value (recall that when a project's return exceeds the cost of capital the NPV is positive).

But in the insurance application, we are using the IRR to determine the underwriting profit margin in the case when the IRR is equal to the cost of capital (the target rate). Therefore, the problem noted above isn't relevant.

Also keep in mind that there is a simple way to resolve the problems with mutually exclusive projects. To do so, we identify a baseline project, A, whose IRR exceeds the cost of capital and then consider the alternative project, B, by measuring the IRR of B's incremental cash flows (positive and negative differences relative to the baseline project) only. If that incremental project has an IRR above the cost of capital, then we would want to do this incremental project. Since we know we want to do project A and now we also want to do this incremental project $B - A$, we choose project B instead of project A.

Question 12. Feldblum identifies an important weakness of the IRR method in the context of regulatory rate filings, which causes him to recommend using the net present value (NPV) approach in rate filings even if the IRR method is used internally. Describe the weakness he identifies.

Solution. Feldblum notes that the IRR method suffers from a "presentation problem". Regulators who see a negative NPV can see clearly that this means the rates are inadequate. But showing them an IRR that is positive (possibly even high or in line with current investment yields) but still below the cost of capital may make it hard for them to see that this too implies inadequate rates.

In addition, remember that the amount of surplus required is a theoretical value. Higher expected losses will lead to higher assumed surplus commitments, which in turn will lead to higher estimates of investment income. This will drive the IRR closer to the investment yield and could make it appear reasonable to regulators even if it is below the cost of capital.

Robbin: Internal Rate of Return (IRR) on Equity Flows

Introduction

The Robbin paper outlines seven methods for determining an *underwriting profit margin* (U) for the pricing of a single line of business on a prospective basis. In this Chapter of the notes, I present ONLY his Internal Rate of Return (IRR) method since it is essentially identical to the Feldblum IRR methodology. The rest of his methods are presented in the next Section of these notes.

IRR on Equity Flows

To use Robbin's IRR method, we track the equity flows that would occur had we set up a company to write a *single* policy. By setting the value of the initial investment equal to the present value of the future equity flows to the hypothetical shareholders (or, equivalently, setting the net present value of all flows to zero) we can calculate an internal rate of return (IRR). Further requiring this IRR to equal a target IRR, we can solve for the underwriting profit margin (U).

The Equity Flows

The logic for determining the equity flows is simple. We assume that at inception there is some amount of "required surplus", which induces a negative flow for the shareholders (to the company). At the end of each period, we assume that the company pays out all of its income to the shareholders, a positive equity flow, but also has to maintain an adequate amount of surplus to meet statutory requirements, which might result in additional equity flows, positive or negative, in that period. That is, any increase in required surplus will result in flows from the shareholders to the company; any decreases in required surplus will result in equity flows from the company to the shareholders.

Robbin specifies the equity flows as:

$$\text{Equity Flow} = \text{Statutory Income} - \text{Change in Statutory Surplus}$$

$$EF = INC - SCHNG$$

Two points are important.

First, notice that this method uses accounting measures of income and changes in balance sheet items, as opposed to the cash flows from the insurance operations. This is tricky because the equity flows between the company and the hypothetical shareholders are actual cash flows. However, it is the accounting rules regarding income recognition and reserve requirements that dictate the magnitude of the equity flows. It doesn't matter that the premium was paid up front, providing positive cash flow to the company, if the accounting rules require the company to hold an unearned premium reserve and thereby tie up the cash within the company.

And second, although presented in terms of statutory surplus, GAAP accounting rules may also be important. In practice, both GAAP accounting and Statutory accounting impose restrictions on the change in surplus and so the equity flows should reflect both sets of restrictions.

Key Formula

It is hard to depict a particular formula for this method, since embedded within all of the calculations of statutory income and changes in surplus requirements there are a host of calculations. Nonetheless, it is critical that you be able to calculate the relevant equity flows and the IRR. The former is easy; the latter requires use of your financial calculator and the IRR function to solve the equation:

$$0 = \sum_{j=0}^{n} \frac{\left[INC_j - SCHNG_j\right]}{(1 + IRR)^j}$$

Notice that the summation index runs from $j = 0$ because all of the flows are assumed to occur at the beginning of each period. That means that INC_0 will reflect any underwriting income or loss at inception (limited usually to the acquisition expenses incurred since no premium is earned yet). Similarly, $SCHNG_0$ will reflect the required surplus at inception.

Numerical Example

Consider the following:

- A company writes \$100 of premium which is paid by the policyholder 50% up-front and 50% in one year.
- All of the premium is earned in the first year.
- The expected Loss and LAE payments are 60% of premium and are paid out as follows: 50% in year 1, 30% in year 2 and 20% in year 3.
- Acquisition expenses are 40% of premium, of which is incurred 80% at inception and 20% at the start of the second period. They are paid, however, according to the pattern: 50%, 30%, 20%.
- The initial premium to surplus ratio is 2.5 and at each year-end the required surplus is 40% of carried loss reserves.
- Invested assets earn a 5% return per annum, pre-tax.

Ignore taxes and calculate the following for the next 4 years: UW Income, Surplus, Invested Assets, Investment Income and the IRR on equity flows.

There is a lot here, but let's start with the underwriting income. The following table shows the incurred and paid premiums, losses and expenses at the start of each period. The underwriting income is just premiums less loss and expenses, all on an earned basis:

TABLE 1. IRR Example — UW Income

	Premium		Loss and LAE		Expenses		
Time	Earned	Paid	Incurred	Paid	Incurred	Paid	UW Income
0	0	50	0	0	32	20	-32
1	100	50	60	30	8	12	32
2	0	0	0	18	0	8	0
3	0	0	0	12	0	0	0
4	0	0	0	0	0	0	0

Now we will work on the balance sheet items.

For the most part these just involve tracking the reserves and receivables, but it also involves determining the required surplus each period and the invested assets. To do this, start with the three reserve accounts, then determine the surplus (in this case 40% of premium at inception, based on P:S = 2.5, and 40% of the loss reserve balance beginning at the end of the first period). We then determine the premium receivables balance and plug for investable assets so that the balance sheet is in balance (i.e. Invested Assets = Total Liab & Surplus *less* Prem Receivable).

The columns in the table below follow this progression:

TABLE 2. IRR Example — Balance Sheet Items

	Reserves				Total	Premium	Invested	Average
Time	Unearned Premium	Loss and LAE	Expenses	Surplus	Liab & Surplus	Receivable	Assets	Assets
0	100	0	12	40.0	152.0	50.0	102.0	0.0
1	0	30	8	12.0	50.0	0.0	50.0	76.0
2	0	12	0	4.8	16.8	0.0	16.8	33.4
3	0	0	0	0.0	0.0	0.0	0.0	8.4
4	0	0	0	0.0	0.0	0.0	0.0	0.0

Next, we determine the investment income each period using the invested assets. In Robbin's calculations he used the *average* invested assets, which approximates the timing of the cash flows reasonably.

TABLE 3. IRR Example — Investment Income on Average Assets

Time	Average Assets	Investment Return	Investment Income
0	0.0	5.0%	0.0
1	76.0	5.0%	3.8
2	33.4	5.0%	1.7
3	8.4	5.0%	0.4
4	0.0	5.0%	0.0

And then finally, we can determine the equity flows as the UW income, plus the investment income, less the change in surplus (recall we are ignoring taxes):

TABLE 4. IRR Example — Equity Flows and IRR

Time	UW Income	Investment Income	Change in Surplus	Equity Flow
0	-32.0	0.0	40.0	-72.0
1	32.0	3.8	-28.0	63.8
2	0.0	1.7	-7.2	8.9
3	0.0	0.4	-4.8	5.2
4	0.0	0.0	0.0	0.0
			IRR	**6.6%**

To calculate the IRR on your CAS approved calculators (this is for the TI BA II Plus, but the others would be similar), you would enter the following:

- Clear the cash flow registry by pressing $\boxed{\text{2nd}}$ $\boxed{\text{CE/C}}$.
- Enter the initial cash flow by pressing $\boxed{\text{CF}}$ and then -72.0 $\boxed{\text{ENTER}}$.
- Enter the cash flow amounts and the number of times each cash flow occurs. In this case, we will use the default frequency of 1 for each cash flow and just enter the amounts. For the next cash flow, enter 63.8 $\boxed{\text{ENTER}}$ and then $\boxed{\downarrow}$ $\boxed{\downarrow}$. The first down arrow accepts the default payment frequency of 1 and the second one moves to the next cash flow registry entry.
- Enter the remaining cashflows as $xx.xx$ $\boxed{\text{ENTER}}$ $\boxed{\downarrow}$ $\boxed{\downarrow}$
- Then compute the IRR with the entries $\boxed{\text{IRR}}$ $\boxed{\text{CPT}}$

Advantages and Disadvantages

An advantage of this approach is that the IRR is relatively simple to interpret — it reflects the expected "interest" earned on the shareholders' "loan" of surplus to the company. It also nicely captures the impact that accounting rules and regulatory capital restrictions directly impact shareholders.

However, to use this we need to assume a particular surplus requirement, at inception and over time, which may not make sense on a line of business or policy basis. We also need to identify a target IRR. And then that forces us to confront the overall challenge that riskier lines may need more surplus, may require a higher IRR or maybe both. Having to set both is therefore tricky.

Feldblum's IRR Paper

Feldblum's IRR paper discusses the exact same method and provides several additional details and considerations. Be sure to review that paper in conjunction with this one.

Practice Questions

Question 1. Some discounted cash flow and IRR models used by actuaries focus solely on the underwriting cash flows of premium, loss and expense. Why does Robbin advocate focusing on cash flows to equity holders rather than just policyholder cash flows.

Solution. The policyholder cash flows don't take into account the impact of statutory or GAAP accounting conventions that impact shareholders, such as the need to carry reserves, to recognize deferred acquisition costs under GAAP, etc. They also don't take into account surplus requirements.

Question 2. You are trying to calculate the IRR for a particular insurance policy using the same approach that Robbin used for his IRR on Equity Flows method. You are given the following assumptions:

- Premium: 100, written at inception and earned entirely in Year 1, collected 75% at inception and 25% in Year 1
- Losses: 72, paid out 25% in Year 1, 50% in Year 2, 25% in Year 3
- Expenses: 30, 60% of which is incurred at inception and 40% incurred in Year 1, and paid 30% at inception , 45% in Year 1 and 20% in Year 2 and 5% in Year 3
- Required Surplus (Statutory): 31.5% of unpaid losses
- Investment Income: 6% per annum on beginning of year invested asset balance
- All cash flows occur at the end of the period

- There are no income taxes
- We care only about Statutory financial statements

Determine the Statutory Underwriting Income at inception and in Years 1 through 3.

Solution. To calculate the Statutory UW Income, we need to track earned premium, incurred loss and incurred expense.

Year	Earned Prem	Incurred Loss	Incurred Expense	Stat UW Income
0	0	0	18	-18
1	100	72	12	16
2	0	0	0	0
3	0	0	0	0

Question 3. Using the same information as in the previous question, determine the surplus required at inception and at the end of Years 1 through 3.

Solution. Based on the specific wording of the information given, required surplus is 31.5% of the unpaid losses. At inception, while no exposures have technically been earned, the unpaid loss is 72. Under normal insurance accounting this would not be set up as a loss reserve, but for the purposes of this question it would be used to say that the required surplus at inception is $31.5\% * 72 = 22.7$.

The rest of the values are calculated similarly, using the loss reserve balances.

Year	Incurred Loss	Paid Loss	Unpaid Loss	Required Surplus
0	0	0	72	22.7
1	72	18	54	17.0
2	0	36	18	5.7
3	0	18	0	0.0

Notice that the calculation of the initial required surplus would have been very different had the question stated that the required surplus was a percentage of the loss reserve or the total reserves. In the case of the loss reserves, there are technically no loss reserves at inception. In the case of the total reserves, there is an unearned premium reserve at inception equal to the written premium.

Question 4. Continuing with the same facts as in the previous questions, determine the total asset values, the invested asset balances and the investment income at inception and at the end of Years 1 through 3. Hint: Remember to assume that all assets not required to be held as reserves or surplus will be paid to investors each period.

Solution. Start with the total assets. Because of the point made in the Hint, we know that each period the total assets are the sum of the required surplus and all of the reserves. The unearned premium and the loss reserves are easy, but don't forget that there are incurred but unpaid expenses each period too. Whether you call this a "reserve" or just a liability, it will need to be included in the calculation. Notice that in order to determine the reserves, I need to pay attention to the incurred loss and expense as well as the paid loss and expense, so let me start by showing the payments:

Year	Prem	Loss	Expense
0	75	0	9.0
1	25	18	13.5
2	0	36	6.0
3	0	18	1.5

Using the incurred and paid premium, loss and expense information, it is easy to then show the reserves, liabilities and surplus to get the total assets:

	Reserves & Liabilities				
Year	UEP	Losses	Expenses	Required Surplus	Total Assets
0	100	0	9.0	22.7	131.7
1	0	54	7.5	17.0	78.5
2	0	18	1.5	5.7	25.2
3	0	0	0.0	0.0	0.0

Finally, we can get the invested assets by just recognizing that all assets are invested except for the amounts receivable (i.e. the unpaid premium balances). The table below shows the invested assets and the associated investment income, which is 6% of the beginning of period assets.

Year	Total Assets	Receivables	Invested Assets	Investment Income
0	131.7	25	106.7	0
1	78.5	0	78.5	6.4
2	25.2	0	25.2	4.7
3	0.0	0	0.0	1.5

Question 5. Using the same information as in the previous questions, determine the total equity flows to investors at inception and in Years 1 through 3 and then calculate the IRR for this policy.

Solution. Investors can take out of the company the UW income, the investment income and the decrease in the required surplus each period. Of course, at inception the UW income is negative and the required surplus increases, so these will result in flows from the investors to the company.

Each period, the amounts will be as follows:

Year	UW Income	Inv Income	Decrease in Surplus	Total Flow to Investors
0	-18.0	0.0	-22.7	-40.7
1	16.0	6.4	5.7	28.1
2	0.0	4.7	11.3	16.1
3	0.0	1.5	5.7	7.2

Finally, we calculate the IRR using the equity flows in the last column, which results in $IRR =$ 16.1%.

Notice that the investors have to contribute equity capital to cover the up-front UW loss so that after recognition of this loss the surplus is 31.5% of the unpaid losses.

Question 6. Suppose we wanted to take into account GAAP accounting in the previous questions. What adjustment does Robbin make and how would it affect the IRR calculated above?

Solution. To reflect GAAP accounting, Robbin takes into account the requirement to capture Deferred Acquisition Costs (DAC). This has the effect of deferring the recognition of the acquisition expenses until the premium is earned. In our case, this would eliminate the $18 of expenses incurred at inception under Statutory accounting and instead incur all expenses in Year 1.

The most important impact that this has is on the underwriting income, which will be higher at inception and lower in Year 1. Robbin assumes that the income taxes are unaffected by this though, since they are calculated based on statutory income *in the paper on the syllabus*.

Robbin makes the additional assumption that the DAC balance is *added to* the statutory surplus to arrive at the required GAAP capital. What this means is that while the income is different, the required capital is also different and there is actually no impact on the equity flows to investors. The IRR is unchanged.

Selected Old Exam Questions for Part 6

The following questions relevant for this section appeared on the Old CAS Exam 9 between 2002 and 2010 and on the New CAS Exam 9 since 2011.

Feldblum IRR	**Robbin IRR**
2004 Q31	2002 Q22
2004 Q35	2007 Q15
2005 Q16	2008 Q14
2005 Q18	2008 Q8
2006 Q14	2009 Q7
2006 Q15	2010 Q13
2006 Q16	
2006 Q21	
2007 Q21	
2007 Q22	
2008 Q9	
2009 Q15	
2009 Q16	
2010 Q11	
2011 Q19	
2012 Q17	
2013 Q20	
2014 Q18	
2015 Q19	
2015 Q20	

For some of these questions I have provided the text of the question and an explanatory solution. These were selected either because they are representative of the questions you are likely to be asked on future exams or because they contain an element that is particularly worthwhile to point out. For the other questions, the CAS solutions should be sufficient to confirm whether your answer is correct.

Note that this section only includes the Internal Rate of Return portion of the Robbin reading. All other methods discussed in that reading are included in the next section of these notes.

Important Note: The solutions shown here are intentionally detailed. They contain thorough explanations of the concepts and formulas used to reinforce the main points from the readings and provide an additional teaching opportunity. **Actual exam responses should be much more concise than what is shown here, along the lines of what you will see in the solutions that the CAS releases.**

2010 Exam Question 7

A monoline insurance company is operating under the following conditions:

Expected Loss	500
Ratio of Undiscounted Reserves to Surplus	2.0
Required Rate of Return	15%
Investment Yield	5%
Risk Free Rate	3%

Additionally, you are told all premium is collected on Jan 1, 2011; all losses are paid on Jan 1, 2012; there are no other expenses or taxes; the risk is independent of all other risks and the risk has a one-year policy term effective Jan 2, 2011.

Note: The question was not clear, but for the purposes of my answer I am assuming that the new risk is just like the old risks, so that it has an expected loss of $500.

Calculate the minimum amount of premium to be charged to the new risk to meet the required return.

There are multiple ways to interpret this question, so here I will just show how it can be done using the IRR methods in Feldblum and Robbin. The next section of the notes will also include solutions to the same question based on the Ferrari reading and the other methods discussed in Robbin.

For the IRR methods in Feldblum and Robbin, you have to track the flows to the equity owners. In this case, when the policy is written they contribute $250 of surplus. At the end of the year they will have flows out of the company equal to the total income earned less the increase in surplus.

The income earned reflects the underwriting profit (the unknown U) plus the investment income, so the flows out are $U + 5\%(500 + 250)$ at the end of the year. On Dec 31, 2011 we still have a 500 reserve, so the surplus remains at 250. Then, on Jan 1, 2012 we pay out the claim and take the surplus down to zero. We don't have any income on Jan 1 and since the difference is just a day, let's just assume that the income and the take down of surplus occur at the same time.

This gives us an IRR calculation of:

$$0 = -250 + \frac{U + 5\%(500 + 250) + 250}{1 + IRR}$$

Setting IRR = 15% and solving for U gives us $U = 0$.

2010 Exam Question 11

The following information applies to an insurance company with a surplus balance of \$100,000.

Line of Business	Written Premium	Exp Loss Ratio
Homeowners	200,000	50%
Workers Compensation	100,000	70%

The average duration from occurrence to payment of a claim is .5 years for homeowners and 4.0 years for workers compensation. Assume the company is in a steady state environment.

a. Assuming surplus is committed when the policy is written and is released when the policy expires, calculate the allocated surplus for each line of business.

Note that they didn't specify the method of allocation, only the timing of its commitment and release. A natural assumption in this case, which is how Feldblum did it in his examples on this topic, is that the surplus is allocated in proportion to written premiums, in which case 2/3 of it would be allocated to the homeowners line and 1/3 to the workers compensation line.

So $\$66,667$ of surplus for homeowners and $\$33,333$ of surplus for workers compensation. These would be allocated at the start of the year and released at the end.

b. Assuming surplus is committed when the losses occur and is released when the losses are paid, calculate the allocated surplus for each line of business.

Here, if we are allocating surplus when losses occur it would be natural to do so in proportion to loss reserves. This, again is what Feldblum did in his example, but the language of the question makes this ambiguous.

Nonetheless, there is one other issue to keep in mind and that is the steady state assumption. The homeowners line has $\$100,000$ of expected losses based on the premium and loss ratio information given and loss reserves are held on average for half a year. In a steady state, the company would always have one half year's worth of loss reserves, or $\$50,000$.

Similarly, the workers compensation line has $\$70,000$ of expected losses each year and the reserves are held for four years. In a steady state, there would be $\$280,000$ of reserves held.

This would result in $50/330 = 15.15\%$ allocated to homeowners ($\$15,150$) and $280/330 = 84.85\%$ allocated to workers compensation ($\$84,850$).

c. Surplus protects the insurer against several types of risk. Identify and briefly describe two sources of risk which occur only during the policy period.

Feldblum listed several risks, including the following:

- Asset risk - loss in value of marketable securities
- **Pricing risk - risk that incurred loss and expense is greater than expected**
- Reserving risk - risk that loss reserves will not cover ultimate loss payments
- Asset-liability mismatch risk - interest rate risk
- **Catastrophe risk - risk of hurricane or earthquake**
- Reinsurance risk - risk that reinsurance recoverables won't be collected
- Credit risk - risk that receivables will not be collected

The two risks in bold, pricing risk and catastrophe risk occur during the policy period, while the others continue until losses are paid.

2010 Exam Question 13

The following information is applicable for a policy rated using the IRR method and annual evaluation points.

Expected loss	750,000
Ratio of expense to premium	22%
Investment yield (pre-tax)	4%
Federal income tax rate on UW income	35%
Effective federal income tax rate on investment income	20%
Date premium is collected and expense paid	1-Jan-11
Date all losses and income taxes are paid	1-Jan-12
Target IRR on equity flows	10%

The appropriate surplus requirement is 30% of the sum of the unearned premium and loss reserves.

Calculate the underwriting profit provision.

This is a simple IRR calculation because all of the losses are paid at the end of one year. There are only two equity flows that are relevant — the initial funds provided up-front to ensure adequate surplus and the expected after-tax net income plus surplus returned at the end of one year.

Let's first note a few points about the necessary calculations:

- The initial surplus that must be held at inception is specified as being 30% of the sum of the unearned premium and the loss reserves. In this case, at time zero all of the premium is unearned and there are no loss reserves, so the initial required surplus is just 30% of the premium.

- An important issue to address is when to incur the expenses (22% of premium). We know they are paid up-front, and on a Statutory basis they would result in an UW loss at inception. In the Robbin paper, this also resulted in a tax benefit at inception, but in this case the question said that all taxes are paid (and presumably credited) at the end of the period. Because of this end-of-period treatment of the taxes, we could just defer the expenses and incur them at the end. But here, I will assume they are incurred at inception and I will be careful about the when the taxes are reflected. I will pay UW income tax on the cumulative UW profit through the end of the year.

- At each year-end we would have to reevaluate the required surplus. In this problem, they said the losses were paid on Jan 1, 2012. To be technically accurate you might argue that this means the company in this example has to hold loss reserves and surplus at the end of 2011 and then can release that surplus the next day once the losses are paid. I will ignore this minor timing issue and just assume all surplus is taken down at the end of the year.

- The investment income earned depends on the assets invested. As described above, total assets at inception are equal to the sum of the unearned premium reserve and the required surplus. Given that all payments occur at the end of the period, the assets are fully invested for the entire year (in the paper Robbin used the average assets during the period).

- Since we need to solve for the premium that achieves the target IRR, and since the initial equity flow, the investment income and the underwriting income all depend on this, we need to keep the algebra straight. Before showing the algebra, consider the following table with the key intermediate calculations (and the correct final solution included):

	Premium		Loss and LAE		Expenses		
Time	Earned	Paid	Incurred	Paid	Incurred	Paid	UW Income
0	0	981,675	0	0	215,969	215,969	-215,969
1	981,675	0	750,000	750,000	0	0	231,675

		Reserves			Total	Premium	Invested
Time	UEP	Loss and LAE	Expenses	Surplus	Liab & Surplus	Receivable	Assets
0	981,675	0	0	294,503	1,276,178	0	1,276,178
1	0	0	0	0	0	0	0

	Pre-tax Income			Decrease in	
Time	UW	Investment	Taxes Paid	Surplus	Equity Flow
0	-215,969	0	0	-294,503	-510,471
1	231,675	51,047	15,707	294,503	561,518

The above table was shown solely to make it clear how the calculations are done. The steps below are the algebra that is really needed.

We can see from the table above that the initial capital contribution is equal to the up-front expenses (22% of Premium) plus the required capital (30% of Premium). The full premium and required capital is invested (the investors paid for the up-front expenses as part of their capital contribution) of the full year.

We need to solve the following for the premium, P:

$$0 = -[.22P + .30P] + \frac{P - 750,000 - .35(P - .22P - 750,000) + [P + .30P](.04)(1 - .20) + .30P}{1.10}$$

Solving for the premium requires several steps, but can be found to be $P = 981,675$.

The underwriting profit provision is then found by calculating the combined ratio and subtracting that from 1.0, or $U = 1 - (L + E)/P = 1.6\%$.

2011 Exam Question 18

An insurance company is analyzing a policy for the upcoming policy period. Given the following information:

- The policy is in force for one year.
- Premium is collected at the beginning of the policy period.
- Losses are paid according to the following schedule: \$40,000 at the end of Year 1, \$35,000 at the end of Year 2 and \$30,000 at the end of Year 3.
- At the beginning of the policy period \$100,000 of capital must be **committed** and will be released entirely at the end of year three.
- A 5% annual return is earned on investment of the capital and reserves.
- Investment income is earned at the end of a given year.
- An administrative expense equal to 15% of the policy premium is incurred and paid at the beginning of the policy period.
- A tax rate of 35% applies to all income. Tax savings due to losses in a given period can be counted as an asset.
- For underwriting income purposes, assume the policy premium is earned in full by the end of year one.

Calculate the premium the insurance company must charge in order to meet its goal of a 15% internal rate of return on this policy.

Important Note: The original question included assumptions about total reserve balances (unearned premium and loss reserves) over time. This language was inconsistent with some of the other assumptions and so I have deleted it and ignored it in my solution. I discuss this further after my version of the solution to the problem as stated above.

The hardest part of questions like this one is to get the accounting and cash flow items properly reflected. But when the question isn't worded very carefully, it is easy to misinterpret the assumptions. This explains why the CAS seems to have accepted 7 different answers to this question.

Let's start with the premium. Whatever amount we charge, call it P, we earn it entirely during the first year and incur a total of $\$105,000$ in losses (claims). We also incur 15% of P as expenses up-front. It is these up-front expenses that create the biggest challenge in problems like this, so I will explore this topic in detail below.

In the Robbin paper, he allowed for there to be incurred expense (negative income) at inception of the policy. The effect of this is that surplus is reduced at inception so that an immediate capital contribution (over and above the stated "required capital") is needed at inception. As an alternative, you could think of these expenses as being deferred (as under GAAP) but having to set up a deferred acquisition expense (DAC) asset that increases the required capital under GAAP. Either way, the investors need to contribute this amount.

There is an unfortunate difference though between the two approaches above. In the paper on the syllabus, Robbin calculated taxes according to the Statutory income, even though GAAP income differed (because of the deferral of expenses). This resulted in an immediate tax benefit in the example in the paper (though, as an aside, the tax rates and the income figures don't reconcile). In subsequent papers and presentations of this method, Robbin has changed the way this is handled and started to use the GAAP income to calculate the taxes, eliminating this up-front tax benefit. For the moment, let's do it the way it was done in the paper that is actually on the syllabus and treat the negative UW income at inception as having an associated tax benefit.

Investment income is earned each period based on the beginning of the period invested asset balances. In Robbin's paper he used the average balances, but here the assumption is that all premiums and expenses are paid up-front and all losses are paid at the end of the period so any funds invested are invested for the full year. I will use the beginning of the year invested asset balance to determine the investment income.

To use the IRR method, as in the Robbin reading, we need to track the cash flows to and from the capital providers. In the question it said that $\$100,000$ is contributed at inception and held for three years. Usually in problems like this one we are told what the capital has to be (as a proportion of written premium or a proportion of reserves) *after* we recognize the up-front expenses. Investors then have to contribute enough to cover these up-front expenses and the required capital. While not what was intended, in my initial solution below I am going to interpret the question in this way and assume that the surplus balance has to be $\$100,000$ at inception, after the tarter-tax UW loss is recognized, and that investors have to contribute those funds as well at inception and then receive them back in Year 1 when the income is recognized. In an alternate solution I will modify this assumption to be more consistent with the wording in the question.

Finally, on the actual exam you have a very hard task because you need to track the after-tax underwriting and investment income as a function of the unknown P and then solve the IRR problem for P. I will get to that in a moment, but for now let's just look at the key calculations given my final answer for $P = 160,758$.

Year	Earned Premium	Incurred Loss	Incurred Expense	Pre-Tax UW Income		Paid Premium	Paid Loss	Paid Expense
0	0	0	25,385	-25,385		169,236	0	25,385
1	169,236	105,000	0	64,236		0	40,000	0
2	0	0	0	0		0	35,000	0
3	0	0	0	0		0	30,000	0

Year	UEPR	Loss Reserve	Total Reserve	Surplus	Total Liab and Surplus	Premium Receivable	Invested Assets	Investment Income
0	169,236	0	169,236	100,000	269,236	0	269,236	
1	0	65,000	65,000	100,000	165,000	0	165,000	13,462
2	0	30,000	30,000	100,000	130,000	0	130,000	8,250
3	0	0	0	0	0	0	0	6,500

Year	UW Income	Investment Income	Total Income	Income Tax	After-Tax Income	Change in Surplus	Equity Flow
0	-25,385	0	-25,385	-8,885	-16,500	100,000	-116,500
1	64,236	13,462	77,698	27,194	50,503	0	50,503
2	0	8,250	8,250	2,888	5,363	0	5,363
3	0	6,500	6,500	2,275	4,225	-100,000	104,225

There are three distinct sections to the above exhibit.

The top section shows the pre-tax UW income and the timing of the various UW cash flows, which together help us to be sure we have the proper balance sheet items.

The middle section shows the reserve balances and surplus balance in the first five columns. All of those amounts follow directly from the assumptions and basic insurance accounting, but as noted above I am assuming that we need the surplus balance to be \$100,000 at inception, after recognizing the after-tax UW loss at inception.

The bottom section completes the calculations. It shows the total income before and after tax. The investors contribute the funds to cover the after-tax loss at inception, the \$100,000 in capital then remove all after-tax income as it is earned (see Alternate Solution 1 below for a different assumption about the timing of the distributions to investors).

The final series of cash flows are then used to solve for the IRR, as follows:

$$0 = -116,500 + \frac{50,503}{1 + IRR} + \frac{5,363}{(1 + IRR)^2} + \frac{104,225}{(1 + IRR)^3}$$

This leads to $IRR = 15\%$.

If that made sense, now let's see what the calculations on the exam would have had to be like in order to solve for the unknown premium given the target IRR of 15%.

Instead of the first term $(-116,500)$ we would have had to write this in terms of the unknown P as

$$\text{Initial Capital Flow} = -100,000 - .15P(1 - .35)$$

Then, instead of $50,503$ as the numerator in the second term, we would have had the after-tax underwriting profit (note we've already recognized the expenses) and investment income:

$$[P - 105,000 + .05(P + 100,000)](.65)$$

The value of P doesn't impact the last two terms (recall that I am ignoring the portion of the question that dealt with the timing of the reserve releases).

We can then easily solve for $P = 169,236$ in the following equation:

$$0 = -100,000 - .15P(1 - .35) + \frac{[P - 105,000 + .05(P + 100,000)](.65)}{1.15} + \frac{5,363}{1.15^2} + \frac{104,225}{1.15^3}$$

Alternate Solution 1

To be consistent with the reading on the syllabus, I found the solution above by assuming that the initial capital contribution had to result in the surplus being $100,000 after recognition of the after-tax UW loss at inception. To be more consistent with the wording given, I could have assumed that the total investor contribution up-front was $100,000, used some of that to pay the after-tax UW loss, and then left the surplus balance smaller at inception. If I did this but continue to allow all after-tax income to be paid to investors as it is earned, then the following modified calculations would result in an initial premium of $160,758.

Year	Earned Premium	Incurred Loss	Incurred Expense	Pre-Tax UW Income		Paid Premium	Paid Loss	Paid Expense
0	0	0	24,114	-24,114		160,758	0	24,114
1	160,758	105,000	0	55,758		0	40,000	0
2	0	0	0	0		0	35,000	0
3	0	0	0	0		0	30,000	0

Year	UEPR	Loss Reserve	Total Reserve	Surplus	Total Liab and Surplus	Premium Receivable	Invested Assets	Investment Income
0	160,758	0	160,758	84,326	245,084	0	245,084	
1	0	65,000	65,000	84,326	149,326	0	149,326	12,254
2	0	30,000	30,000	84,326	114,326	0	114,326	7,466
3	0	0	0	0	0	0	0	5,716

Year	UW Income	Investment Income	Total Income	Income Tax	After-Tax Income	Change in Surplus	Equity Flow
0	-24,114	0	-24,114	-8,440	-15,674	84,326	-100,000
1	55,758	12,254	68,012	23,804	44,208	0	44,208
2	0	7,466	7,466	2,613	4,853	0	4,853
3	0	5,716	5,716	2,001	3,716	-84,326	88,042

Alternate Solution 2

Alternate Solution 1 still didn't entirely match the wording of the question, since the question said that investors would only get their $100,000 initial capital contribution back at the end of Year 3. I assumed that, consistent with the syllabus reading, that all after-tax UW income was paid to investors as it was earned.

To account for the specific wording in the question, the most reliable way to modify my solution and ensure that the proper investment income is earned on the retained funds is to assume that at the end of Year 1 the surplus balance has to return to $100,000. That is, I allow the surplus to dip when the up-front expenses are incurred but then when the premium is earned I don't allow it to be paid to investors yet.

The revised calculations require an initial premium of $165,904 in order to achieve a 15% IRR.

Year	Earned Premium	Incurred Loss	Incurred Expense	Pre-Tax UW Income		Paid Premium	Paid Loss	Paid Expense
0	0	0	24,886	-24,886		165,904	0	24,886
1	165,904	105,000	0	60,904		0	40,000	0
2	0	0	0	0		0	35,000	0
3	0	0	0	0		0	30,000	0

Year	UEPR	Loss Reserve	Total Reserve	Surplus	Total Liab and Surplus	Premium Receivable	Invested Assets	Investment Income
0	165,904	0	165,904	83,824	249,728	0	249,728	
1	0	65,000	65,000	100,000	165,000	0	165,000	12,486
2	0	30,000	30,000	100,000	130,000	0	130,000	8,250
3	0	0	0	0	0	0	0	6,500

Year	UW Income	Investment Income	Total Income	Income Tax	After-Tax Income	Change in Surplus	Equity Flow
0	-24,886	0	-24,886	-8,710	-16,176	83,824	-100,000
1	60,904	12,486	73,390	25,686	47,703	16,176	31,528
2	0	8,250	8,250	2,888	5,363	0	5,363
3	0	6,500	6,500	2,275	4,225	-100,000	104,225

Alternate Solution 3

I mentioned in my first solution that the original problem on the exam had different language about the reserve balances. It said to assume that the total reserve balances (unearned premium plus loss reserves) were 100% of the premium amount through year 1, then 50% through year two and then 25% through year three. Not only does this complicate the calculation of the underwriting income each period and introduce yet another term with P in the calculations, it also seems oddly inconsistent with the other assumptions regarding the payment of claims and the earning of the premium. I chose to ignore that in the answer above but all of the CAS solutions attempted to take this into account.

Here is a version of the calculations that is more consistent with this aspect of the original question, using the assumption that the expenses are incurred up-front and that investors contribute enough capital to cover the after-tax loss at inception plus leave $100,000 in the surplus account:

Year	Earned Premium	Incurred Loss	Incurred Expense	Pre-Tax UW Income		Paid Premium	Paid Loss	Paid Expense
0	0	0	25,871	-25,871		172,470	0	0
1	172,470	126,235	0	46,235		0	40,000	0
2	0	-8,118	0	8,118		0	35,000	0
3	0	-13,118	0	13,118		0	30,000	0

Year	UEPR	Loss Reserve	Total Reserve	Surplus	Total Liab and Surplus	Premium Receivable	Deferred Expenses	Invested Assets
0	172,470	0	172,470	100,000	272,470	0	0	272,470
1	0	86,235	86,235	100,000	186,235	0	0	186,235
2	0	43,118	43,118	100,000	143,118	0	0	143,118
3	0	0	0	0	0	0	0	0

Year	UW Income	Investment Income	Total Income	Income Tax	After-Tax Income	Change in Surplus	Equity Flow
0	-25,871	0	-25,871	-9,055	-16,816	100,000	-116,816
1	46,235	13,624	59,859	20,951	38,908	0	38,908
2	8,118	9,312	17,429	6,100	11,329	0	11,329
3	13,118	7,156	20,274	7,096	13,178	-100,000	113,178

Here, we had a more complicated calculation to determine the UW income or loss each period, since we had to capture the incurred loss each period as the paid loss plus the change in loss reserves. Since we only had to maintain $100,000 in the capital account, we were able to distribute different amounts to investors and there was a modest impact on the investment income as well. But otherwise the calculations were the same.

2011 Exam Question 19

An insurance company writes two lines of business and wants to determine which line is more profitable. Given the following information for the two lines of business:

	Line A	Line B
Premium	200,000	100,000
Reserves	125,000	90,000
Expected Loss Ratio	70%	75%
Expense Ratio	25%	20%
Avg Time to Payment of Claims (Years)	2	4.5

- The policy period is for one year.
- All premium is collected at the beginning of the policy period.
- All expenses are paid at the beginning of the policy period.
- The average time of claim payment is relative to the beginning of the policy period.
- The reserves include both loss and unearned premium reserves.
- The total amount of capital held by the insurance company is $550,000.
- Ignore investment income and taxes.

a. Calculate the IRR for Line A if capital is allocated by line of business based on reserves.

This question was controversial immediately after the exam, with intense debate regarding what was intended. It turns out that even with the benefit of the CAS sample solution, it's hard to know what was intended. I will provide what I think is the right way to think about this problem, though I am not confident my answer would have received full credit

In Section III of the reading, Feldblum discussed the difference between allocating capital in proportion to premiums and allocating capital in proportion to reserves. He argued that in the case of allocating in proportion to premium, it is common to assume the capital is committed when the policy is written and released when the policy term expires. He also argued that if allocation is done using reserves (loss reserves plus the loss portion of unearned premium reserves) it is common to assume capital is committed when the policy is written and released when the claims are paid.

He then went on to provide a numerical example of two lines of business and discussed what would happen in a steady state. This steady state issue is important, but the sample CAS solution ignores this.

More importantly, Feldblum never used a steady state allocation of actual surplus in an IRR calculation. Instead, he always assumed the surplus required was a stated multiple of premiums, reserves or both. This is an important point because in the IRR calculation, as Feldblum argues in the paper, we have to explicitly account for the fact that capital will be held for multiple periods. If we took that issue into account in the determination of the allocation, then we would in some senses be double counting. We would be penalizing the new business for the fact that we are still holding capital to support the run-off of the old business. In other words, just because Feldblum showed what the overall effect would be in a steady state, that doesn't necessarily mean that the IRR on prospective business should be evaluated using that amount.

The only sensible way to evaluate the profitability of a given line of business is to evaluate it *prospectively* with respect to future business being written.

Let's start with Feldblum's steady state analysis. In a steady state (same premiums and loss ratios every year) the reserves for Line A will be equal to two times the expected loss or $2 * 200,000 * 70\% = 280,000$. Similarly, the Line B reserves will on average be $4.5 * 100,000 * 75\% = 337,500$. If we allocate the full $\$550,000$ in surplus in proportion to these steady state reserves, we will allocate $249,393$ to Line A.

But consistent with the argument I made above, some of this capital supports the run-off of old business and some supports the new business. Although not addressed in the paper, I think it makes sense to say that half of this supports the old Line A business and half supports the new Line A business. In other words, if I had to calculate the IRR on the **new** Line A business being written, I would only include half of this amount, or $124,696$.

What did the CAS do in their sample solution? They used the specified reserves, which seems incorrect to me. First, it isn't possible to know if those are existing reserves, but they aren't large enough to reflect any estimate of future reserves based on the premium and loss ratio assumptions. Second, as stated above, it isn't possible to evaluate future business based on capital allocated to past business.

IRR Calculation

We aren't given enough information to know how the reserves are paid out, so I will make a reasonable assumption that they are all paid exactly two years after inception (in one bullet payment). If the reserves don't decline during the period then we can also make the assumption that the capital allocation doesn't either. Together, these simplify the calculations.

Investors contribute the allocated capital at inception. They collect $150,000$ of net premium and have to hold $140,000$ in loss reserves until year 2 when they pay the claims and release the capital that was allocated. I will also assume that they can pay out the $10,000$ in underwriting income at the end of the policy period after the premium is fully earned.

The IRR is found by solving the following equation which focuses on all of the cash flows to and from the investors:

$$0 = -124,696 + \frac{10,000}{1 + IRR} + \frac{124,696}{(1 + IRR)^2}$$

Solve for $IRR = 4.1\%$.

Notice that I doubt this is what was expected. We know from the sample solution that they accepted an answer that allocated capital in proportion to the existing reserves, and that's clearly wrong. It's more likely they intended you to allocate the capital based on the full steady state reserves. But I'd rather give you what I think is the right way to think about this.

b. Calculate the IRR for Line A if capital is allocated by line of business based on premiums.

In this case, it's far less ambiguous what was intended for the capital allocation. The allocation of the company's total surplus to Line A would be $200,000/300,000(550,000) = 366,667$. Using Feldblum's arguments, the allocation is done based on premium it is common to assume it is released when the policy period ends (as opposed to when claims are paid as in Part (a)). The IRR is found by solving the following equation:

$$0 = -366,667 + \frac{366,667 + 10,000}{1 + IRR}$$

Solve for $IRR = 2.73\%$.

But notice something problematic here. In Part (a) I tried to be careful and not evaluate the prospective business based on the capital allocated to older business. Now it is hard to do the same thing. Because I am proportionately allocating the total surplus, it's hard to know what portion of that is really there to support the new business rather than the old. We'd have to assume that the capital always supports just the new business, which isn't realistic. But it is consistent with assuming that the capital is only allocated during the policy period.

c. Briefly describe two reasons why the premium based allocation scenario and the reserves based allocation scenario may provide different answers regarding which line has a higher internal rate of return.

One reason is that when reserves are used, the payment pattern for the claims impacts the relative allocation of initial surplus.

Another reason is that when the allocation is based on premium it is common to assume that the capital is released at the end of the policy term rather than as claims are paid.

So both the amount of capital and the timing of its expected release vary.

2012 Exam Question 17

A P&C insurance company is doing a rate review for one of the lines of business that it writes. Given the following information for this line of business:

- Policies are in-force for one year
- Premium is fully earned at the beginning of the policy period
- Premium is fully collected at the beginning of the policy period
- A fixed expense amount of $50 per policy is paid by the company at the beginning of the policy period
- Variable expenses equal to 20% of the premium are paid by the company three months after the policy effective date
- Losses averaging 35% of the policy premium are paid by the company six months after the policy effective date
- Losses averaging 25% of the policy premium are paid by the company eighteen months after the policy effective date
- The company allocates capital in a way that results in a premium to capital ratio of 2:1 for this line of business
- The company's investment portfolio earns a 5% rate of return per year
- Investment income is earned at the end of a given policy year
- Supporting capital is released twenty-four months after the policy effective date
- Only supporting capital is invested; investment income is not earned on any of the other assets
- Ignore taxes

a. Calculate the premium that the company must charge in order to earn a 10% internal rate of return (IRR) for this line of business.

Note that I added the assumption that the premium is fully earned at the beginning of the policy period, which didn't appear in the original exam question. Since no assumption about the timing of the earning of the premium was given, I went ahead and added this to make the solution easy. However, this is not realistic – we wouldn't usually earn the premium, incur expenses and incur losses fully at inception.

But note that by earning the premium at inception, we can immediately distribute the underwriting profit to investors. Since only the supporting capital earns investment income and the supporting capital is not tied to loss or unearned premium reserves, none of the other cash flows really matter besides the initial capital flow, the investment income on the surplus and the return of the supporting capital.

I'll start by showing the equity cash flows using the final premium just to make it clear what is going on and then I will show you how to solve for this premium.

	Premium		Loss and LAE		Expenses		
Time	Earned	Paid	Incurred	Paid	Incurred	Paid	UW Income
0.00	319	319	192	0	114	50	14
0.25	0	0	0	0	0	64	0
0.50	0	0	0	112	0	0	0
0.75	0	0	0	0	0	0	0
1.00	0	0	0	0	0	0	0
1.25	0	0	0	0	0	0	0
1.50	0	0	0	80	0	0	0
1.75	0	0	0	0	0	0	0
2.00	0	0	0	0	0	0	0

		Reserves			Total	Premium	Invested	Other	Total
Time	UEP	Loss and LAE	Expenses	Surplus	Liab & Surplus	Receivable	Assets	Assets	Assets
0.00	0	192	64	160	415	0	160	255	415
0.25	0	192	0	160	351	0	160	192	351
0.50	0	80	0	160	239	0	160	80	239
0.75	0	80	0	160	239	0	160	80	239
1.00	0	80	0	160	239	0	160	80	239
1.25	0	80	0	160	239	0	160	80	239
1.50	0	0	0	160	160	0	160	0	160
1.75	0	0	0	160	160	0	160	0	160
2.00	0	0	0	0	0	0	0	0	0

		Investment	Change in			
Time	UW Income	Income	Surplus	Equity Flow	PV Factor	PV(Equity Flow)
0.00	14	0	160	-146	1.00	-146
0.25	0	0	0	0	0.98	0
0.50	0	0	0	0	0.95	0
0.75	0	0	0	0	0.93	0
1.00	0	8	0	8	0.91	7
1.25	0	0	0	0	0.89	0
1.50	0	0	0	0	0.87	0
1.75	0	0	0	0	0.85	0
2.00	0	8	-160	168	0.83	139
					NPV	**0.00**

Notice that when all of the equity flows, which really only occur at $T = 0$, $T = 1$ and $T = 2$ are discounted at 10% per year, the net present value is zero. That is another way of saying the IRR is 10%. If you wanted to actually calculate the IRR, you just need to be careful because these are quarterly cash flows and so the 10% IRR per year is really $1.10^{1/4} - 1 = 2.41\%$ per quarter.

Now let's see how to solve for the premium on the exam. I'll recreate the table, but use the unknown P:

Time	Premium Earned	Premium Paid	Loss and LAE Incurred	Loss and LAE Paid	Expenses Incurred	Expenses Paid	UW Income
0.00	P	P	.6P	0	50+.2P	50	P-.6P-(50+.2P)
0.25	0	0	0	0	0	.2P	0
0.50	0	0	0	.35P	0	0	0
0.75	0	0	0	0	0	0	0
1.00	0	0	0	0	0	0	0
1.25	0	0	0	0	0	0	0
1.50	0	0	0	.25P	0	0	0
1.75	0	0	0	0	0	0	0
2.00	0	0	0	0	0	0	0

Time	UEP	Reserves Loss and LAE	Expenses	Surplus	Total Liab & Surplus	Premium Receivable	Invested Assets	Other Assets	Total Assets
0.00	0	.6P	.2P	.5P	.6P+.2P+.5P	0	.5P	.6P+.2P	.6P+.2P+.5P
0.25	0	.6P	0	.5P	.6P+.5P	0	.5P	.6P	.6P+.5P
0.50	0	.25P	0	.5P	.25P+.5P	0	.5P	.25P	.25P+.5P
0.75	0	.25P	0	.5P	.25P+.5P	0	.5P	.25P	.25P+.5P
1.00	0	.25P	0	.5P	.25P+.5P	0	.5P	.25P	.25P+.5P
1.25	0	.25P	0	.5P	.25P+.5P	0	.5P	.25P	.25P+.5P
1.50	0	0	0	.5P	.5P	0	.5P	0	.5P
1.75	0	0	0	.5P	.5P	0	.5P	0	.5P
2.00	0	0	0	0	0	0	0	0	0

Time	UW Income	Investment Income	Change in Surplus	Equity Flow	PV Factor	PV(Equity Flow)
0.00	P-.6P-(50+.2P)	0	.5P	P-.6P-(50+.2P)-.5P	1.00	P-.6P-(50+.2P)-.5P
0.25	0	0	0	0	0.98	0
0.50	0	0	0	0	0.95	0
0.75	0	0	0	0	0.93	0
1.00	0	.5P(.05)	0	.5P(.05)	0.91	.909(.5P(.05))
1.25	0	0	0	0	0.89	0
1.50	0	0	0	0	0.87	0
1.75	0	0	0	0	0.85	0
2.00	0	.5P(.05)	-.5P	.5P(.05)+.5P	0.83	.826(.5P(.05)+.5P)

As before, we just have to set the NPV to zero and solve for P. Simplifying the values in the last column we have $.1566P - 50 = 0$ which gives us $P = 319.26$.

Commentary About CAS Model Solutions

The CAS Examiner's Report shows three model solutions. Their second solution is identical to the one I showed above, though keep in mind that this approach required me to assume that the premium, losses and expenses were all incurred at inception and that the investors could extract the underwriting profit at inception. Their third model solution is very similar, though in that solution they only recognized the variable expenses when paid. This complicated the calculations because it required cash flows between the company and the investors in the first quarter to cover these expenses.

But what I want to point out is how strange (and incorrect) their first model solution is. In that solution, they assumed that investors could extract the full premium paid, less the up-front expenses, at inception and then just have to pay the expenses and claims as they came due. In other words, they earned the premium up-front and the fixed expenses, but not the variable expense or claims. This is a bizarre model solution because it directly contradicts the point that the Feldblum IRR reading makes. This essentially is taking a "product market" view of the cash flows between the policyholder and the company, rather than reflecting the cash flows between the company and the investors. It's quite strange that it was shown as a model solution on this exam, especially since "not holding reserves" was one of the common mistakes cited in the report.

b. Assume that regulators insist that the company charge premium rates 5% below what was calculated in Part (a). Propose two modifications that the company could make to still earn a target 10% IRR and calculate the revised premium.

I won't show the calculations, but one possible modification would be to distribute capital more quickly. You could also have assumed lower expenses or lower required capital.

Interestingly, the CAS expected you to actually take the time to show the revised premium and wanted you to show that it was indeed at least 5% lower than in Part A, but they didn't want you to have to solve so that your answer resulted in exactly 5% lower premium. This strikes me as a very odd question to ask on an exam, making you redo an identical problem three times to get full credit but then not being clear enough to avoid some people from mistakenly making it a much harder algebra problem than it had to be (and therefore wasting a lot of valuable time).

2013 Exam Question 19

Actuaries analyzing an annual workers' compensation policy have estimated the following:

Expected Losses	1,200
Expense Ratio	35%
Risk-free interest rate	5%
Investment Return for the company	11%
Overall market return	9%
Cost of capital	8%

In addition, the actuaries assume the following:

- Losses will be paid according to the following pattern: $420 at the end of year 1, $780 at the end of year 3
- A loss reserve is set up on the effective date of the policy and decreases as losses are paid
- Expenses are paid on the effective date of the policy
- A tax rate of 35% applies to investment income
- There is no underwriting income tax
- The company carries surplus equal to 75% of undiscounted loss reserves
- Equity flows occur annually

Calculate the minimum premium necessary for this policy to be profitable.

The wording of the question was vague, since there are many ways to define "profitable". Nevertheless, it is easy to surmise from the focus on the surplus, equity flows and cost of capital that the intent was to determine the premium such that the IRR on the equity flows was equal to the cost of capital of 8%.

Let's try to just use common sense to get to this answer as easily as possible.

Start by recognizing that the required surplus each period is equal to 75% of the carried loss reserves. While it is often unclear when the questions intend for you to establish a loss reserve, here they say that the reserve is established at inception, so the reserve balances and the required surplus are given as follows:

	Inception	Year 1	Year 2	Year 3
Loss Reserve	1,200	780	780	0
Required Surplus	900	585	585	0

Now we need to determine how the equity flows from the investors compare to those surplus requirements.

We know that they need to start with 900 in surplus, but that isn't necessarily the equity flow from investors because some of this surplus might come from the UW profit at inception, if there is any. Similarly, if there isn't an underwriting profit, the investors might have to contribute some capital in addition to the 900. Since the goal here is to solve for the initial premium, P, let's operate under the assumption that we will solve for the initial investor capital contribution as follows:

$$\text{Initial Equity Flow} = 900 - [P - .35P - 1200] = 2100 - .65P$$

What that equation says is that investors have to contribute 900 *less* whatever underwriting profit there is after paying expenses and establishing the loss reserve. If there is an underwriting loss, the initial equity flow will be greater than the 900 required surplus.

The only other item that can impact the equity flows to the investors then is the after-tax investment income earned on the combination of the loss reserves and the surplus (the expenses were assumed to be paid out already and the investors were assumed to take any of the underwriting income at inception). The following calculations show the after-tax investment income, which is earned at a rate of 11% per annum on the assets held at the *beginning* of the year:

	Inception	Year 1	Year 2	Year 3
Loss Reserve	1,200	780	780	0
Required Surplus	900	585	585	0
Total Invested Assets	2,100	1,365	1,365	0
Investment Income		150	98	98

Given this, we can now summarize the equity flows using the initial flow from above, the change in surplus each period plus the investment income each period:

	Inception	Year 1	Year 2	Year 3
Equity Flows	$-(2100 - .65P)$	465	98	683

The final step then is to solve for P such that the IRR is equal to 8%. That is, when the flows are discounted at the cost of capital rate of 8% per annum, the present value of the equity flows in years 1 through 3 is equal to the initial investment (another way of saying the net present value of all of the flows is zero):

$$0 = -(2100 - .65P) + \frac{465}{1.08} + \frac{98}{1.08^2} + \frac{683}{1.08^3}$$

$$P = 1,606$$

Notice that I wouldn't normally distribute UW profit at inception. Instead, I usually assume that up-front expenses are paid at inception but that all premium and losses are incurred later. But in this question they said to establish loss reserves at inception, so it is only reasonable to also earn the premium at inception.

2009 Exam Question 7

An insurance company writes auto insurance in a single state. The company's actuaries have estimated the following parameters for the frequency and severity distributions:

	Mean	Coeff. Of Variation
Frequency	2.00%	2.00
Severity	20,000	5.00

Assume the following:

- The insurance company writes only annual policies.
- All policies are effective Jan 1.
- All expenses are paid Jan 1. *No expense information was given though, so I will assume it is zero.*
- All losses are fully paid at the end of the policy term (December 31 of the same year).
- Each insured has an expected loss ratio of 80%.
- For the purposes of investment income, the premium is fully earned on Jan 1. *I think they mean assume that it is fully collected so that it can all be invested.*
- Loss reserves are set up at policy inception and equal the total expected loss. *I think this means that there is no unearned premium reserve, consistent with the previous statement about all premium being earned on Jan 1.*

The actuary is also aware of the following additional information:

- It is required that the company's surplus be equal to 1 standard deviation of the total expected losses.
- The company is expected to earn 2% annual return on investments.
- All insureds are independent of each other.

The variance of pure premium (Y) is given by the following formula:

$$\mathrm{Var}(Y) = E(N)\mathrm{Var}(X) + \mathrm{Var}(N)E(X)^2$$

where N is the frequency and X is the severity.

Calculate the number of policies the insurer would need to write in order to achieve a 10% expected IRR.

This is a simple implementation of the IRR calculation, with the only wrinkle being that you need to write out the IRR in terms of the standard deviation of the losses, solve for this standard deviation and then solve for the number of policies needed to achieve that standard deviation.

The IRR method begins by determining how much capital has to be committed by shareholders. In this case, surplus has to equal one standard deviation of the portfolio losses, which we can write in terms of the number of policyholders, n.

To keep it easy to follow, let's start with the formula given for the variance of the loss for each policy:

$$\text{Variance} = (2\%)((20,000)(5.0))^2 + (2\%(2))^2(20,000)^2 = 200,640,000$$

Since the policies are independent, the portfolio mean and standard deviation are written in terms of the number of policies, n, as:

$$\text{Portfolio Mean} = 400n$$

$$\text{Portfolio Std. Dev.} = \sqrt{n(200,640,000)} = 14,165\sqrt{n}$$

Note that I am assuming there are no expenses because no expense information was provided. Later I will vary this assumption and show that it makes a difference.

So let's determine how much capital has to be committed by shareholders. On Jan 1 we fully earn the premium, pay no expenses and establish a loss reserve equal to only 80% of the premium. Since the remaining 20% of the premium will be recognized as income at inception, it becomes part of surplus and so the capital we need from shareholders is only one standard deviation less 20% of the premium. The expected loss ratio for the book (and for each insured) is 80%. This means that for n policies written we have premium equal to $P = 400n/.8 = 500n$.

So,

$$\text{Initial Capital} = 14,165\sqrt{n} - 100n$$

The next step is usually to determine the income and surplus that can be distributed each year. In the simple one-period cases like this, the process is easier because we get back all of the invested funds, with interest, less the paid loss. We don't have to distinguish between underwriting income, investment income or return of surplus in this case. But I will write it in terms of UW income (premium less loss or 20% of premium in this case) plus the investment income on the net premium and the initial capital provided just for clarity.

Setting the IRR equal to 10% gives us the following to solve for n:

$$\begin{aligned} 0 = &-[14,165\sqrt{n} - 100n] \\ &+ \frac{20\%(500n) + 2\%[500n + (14,165\sqrt{n} - 100n)] + [14,165\sqrt{n} - 100n]}{1.1} \end{aligned}$$

Solving for n gives us $n = 92$.

Because I assumed there were no expenses, I sort of made the algebra harder. A simpler assumption would have been that expenses were 20% of premium.

In this case, the premium fully funds the expenses and the loss reserve and the committed capital is just equal to the surplus requirement. Then the only income is the investment income on the net premium and the surplus, so the above formula with 20% expenses would have been:

$$0 = -14,165\sqrt{n} + \frac{2\%[400n + 14,165\sqrt{n}] + 14,165\sqrt{n}}{1.1}$$

Solving for n gives us $n = 20,064$.

Notice that in either case the investment income is the same (it's the loss reserve and the surplus) but the initial investment term is simpler and the algebra is really easy.

Letting $\alpha = 14,165\sqrt{n}$ we can write the solution as:

$$\begin{aligned}
0 &= -\alpha + \frac{2\%[400n + \alpha] + \alpha}{1.1} \\
1.1\alpha &= 8n + 1.02\alpha \\
.08\alpha &= 8n \\
\alpha &= 100n \\
14,165\sqrt{n} &= 100n \\
\sqrt{n} &= 141.65 \\
n &= 20,064
\end{aligned}$$

Notice that I worked it out the hardest way first and then showed an easier way. I did this on purpose because I think doing it in that order emphasizes how important it is to get the initial investment by shareholders right to keep the balance sheet in balance. Different problems will always have this step, but sometimes it will be harder than in other cases. Either way, I think it is helpful to see that life is easier when the premium funds the expenses and the loss reserve and investors only have to put up the surplus. But in general, you won't really have a choice. We just had one here because the question erroneously excluded the expense information.

2015 Exam Question 19

The actuarial department of a property and casualty insurer is reviewing two potential growth opportunities, referred to as Option A and Option B. Given the following information about these options:

- Policies are in-force for one year
- Premium is collected at the beginning of the policy period
- Variable expense equal to 25% of the premium are paid at policy inception
- Both options are expected to achieve an 85% loss ratio
- Equal loss amounts are expected to be paid out at the end of each year over a two-year period for both options
- The company allocates capital in a way that results in a reserve-to-surplus ratio of 2:1 for Option A and 3:1 for Option B
- Initial required surplus is set to the loss portion of the unearned premium reserve
- Supporting capital (and other assets) can be invested at a 10% rate of return
- The cost of capital is 8% per year
- Investment income is earned at the end of the given policy year
- There are no taxes

Calculate the internal rate of return (IRR) and net present value (NPV) for each growth opportunity.

Although not clearly stated in the question, I assume they wanted the IRR and NPV of the *equity flows*. I will begin by determining all of those equity flows for Option A using the same organized process used for all of the IRR problems. But note that here I am not going to have to solve for the premium, so I am going to just assume the premium is $100 to keep the numbers simple.

Following all of the assumptions specified, I get the following:

Time	Premium Earned	Premium Paid	Loss and LAE Incurred	Loss and LAE Paid	Expenses Incurred	Expenses Paid	UW Income
0	0.0	100.0	0.0	0.0	25.0	25.0	-25.0
1	100.0	0.0	85.0	42.5	0.0	0.0	15.0
2	0.0	0.0	0.0	42.5	0.0	0.0	0.0

Time	UEP	Reserves Loss and LAE	Expenses	Surplus	Total Liab & Surplus	Premium Receivable	Invested Assets
0	100.0	0.0	0.0	85.0	185.0	0.0	185.0
1	0.0	42.5	0.0	21.3	63.8	0.0	63.8
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Time	Pre-tax Income UW	Pre-tax Income Investment	Taxes Paid	Decrease in Surplus	Equity Flow
0	-25.0	0.0	0.0	-85.0	-110.0
1	15.0	18.5	0.0	63.8	97.3
2	0.0	6.4	0.0	21.3	27.6

Notice that the original question somewhat implied that investment income was earned only on supporting surplus, but I earned it on all invested assets as of the end of the period. As always in these problems, the need to pay expenses up-front but hold an unearned premium reserve results in initial capital contributions from investors that are larger than the required surplus ($110 vs. $85).

Using those equity flows, I can calculate the NPV as $3.73, or 3.73% of premium, for Option A by simply discounting each equity flow at the 8% cost of capital rate. I can then set the NPV equal to zero and solve for the IRR, which is 11.03%.

Option B is identical, but for a change in the surplus requirement. Changing the calculations to reflect a 3:1 reserve to surplus ratio, the NPV becomes $3.61, or 3.61% of premium, and the IRR becomes 11.08%.

Part 7

Rate of Return and Risk Loads

Part 7 Table of Contents

Ferrari: Total Return on Owners' Equity

Introduction

This paper was written in response to an analysis of industrywide insurance profitability relative to the total assets held by insurers. The author shows two illuminating breakdowns of the total return to shareholders into its key drivers and then suggests how insurers can use the insights provided to, for instance, establish target levels of leverage that maximize the value of the company from the shareholders' perspective.

An important warning regarding this attempt at optimization is provided by one of the reviewers, but nonetheless Ferrari's breakdown is helpful in understanding what drives total shareholder returns and in showing why actuaries need to broaden their focus beyond underwriting profit only and begin to consider investment returns, leverage and other factors that drive total shareholder returns.

The Notation

The following notation is used:

- T - Total after-tax return to the insurer
- I - Investment gain or loss (after appropriate tax charges)
- U - Underwriting profit or loss (after appropriate tax charges)
- P - Premium income
- A - Total assets
- R - Reserves and other liabilities (excluding equity in unearned premium reserves)
- S - Stockholders' equity (capital, surplus, and equity in unearned premium reserve)

The Basic Equation

Using the above notation, the total return on equity is given by the ratio of T to S. However, if we use the following two basic relationships:

$$T = I + U$$

$$A = R + S$$

then with a few steps of algebra this can be written as:

$$\frac{T}{S} = \frac{I}{A}\left(1 + \frac{R}{S}\right) + \frac{U}{P}\frac{P}{S}$$

What this says is that the total return is driven by both the investment return on assets and the underwriting profitability relative to premium and in each case the quantities are adjusted by leverage factors. In the case of return on assets, we use an insurance leverage factor that considers the size of the reserves relative to surplus. In the case of the underwriting profit we use an insurance exposure factor that considers the amount of premium that can be written relative to surplus.

Relationship to the Industry Profitability Study

Notice that when people talk about insurance profitability, they sometimes take the perspective of the regulator or the pricing actuary and are concerned with profit relative to the premium charged.

In the industry study alluded to earlier, they took the perspective of society overall and were concerned with whether the profitability relative to total assets dedicated to the insurance industry.

And others are more concerned with whether the shareholders of insurance companies, the owners, are earning adequate returns on their investment in the industry. They assess profits relative to the value of the shareholders' equity.

What is nice about Ferrari's basic formula is that it encompasses all three of these different profitability measures and shows how the two forms of leverage impact the resulting total return on shareholders' equity.

Reserves as a Form of Capital and Their Impact

If we take the previous equation and make a few algebraic substitutions, we can rewrite the previous formula for total shareholders' return as:

$$\frac{T}{S} = \frac{I}{A} + \frac{R}{S}\left(\frac{I}{A} + \frac{U}{R}\right)$$

This form of the equation is very helpful in thinking about the role of the reserves as another source of investable capital. Unlike the shareholders' equity, the reserves do not (yet) represent the assets of the owners and instead appear to "belong" to the policyholders. So in a sense, the existence of reserve capital on the balance sheet represents a form of leverage where capital that can be invested in marketable securities to earn the rate I/A is borrowed from policyholders. The implicit cost of this borrowing is the underwriting profit or loss (this time in relationship to the reserves though and not the premium).

One of the key points of this form of the formula is that it shows that total shareholders' return can in fact increase even if the underwriting profit is negative, so long as the underwriting loss is more than offset by investment returns.

You may recognize this argument from Warren Buffet's typical description of the insurance business as a source of *float* that is used to finance his investment activities, where the cost of the float is the underwriting loss.

Debt Leverage and Firm Value

For non-insurance companies explicit debt issuance — borrowing money by issuing bonds — is often used to leverage the return to shareholders. In that case though, the cost of this borrowing is usually fixed, whereas in the case of the implicit borrowing in insurance the cost is uncertain and could at times be quite large. Nonetheless, the concept is similar.

Two points are worth noting about the conventional view of debt leverage. First, the existence of debt (implicit or explicit) serves to make the shareholders' equity more volatile, hence riskier. Second, under certain conditions, while the existence of debt leverage increase the expected shareholders' return it does not increase the value of the shareholders' equity holdings because the extra risk requires that the higher levels of profit be discounted more heavily.

Optimal Capital Structure

The term *capital structure* is generally used to refer to the degree of debt leverage on a company's balance sheet. In the case of insurers, this is equivalent to the ratio of reserves to surplus.

One use of the relationship shown between the total shareholders' equity and the reserve to surplus ratio is that it can be used to help think about the optimal reserve to surplus ratio that leads to a maximization of the shareholders' equity. The challenge though, mentioned previously, is that as the amount of leverage increases the higher rate of return and hence higher earnings are offset by an increase in the riskiness of these cash flows and a decrease in their value. Finding the right balance is therefore trickier than simply maximizing the ratio of T/S.

One important complication is the fact that the more investment risk that is assumed, the lower the prudent reserve to surplus ratio, for instance to achieve other objectives related, for instance, to the probability of insolvency or other considerations.

Balcarek's Review Comments

Balcarek began his review by highlighting a point that was made, although more subtly, by Ferrari that steps to increase total shareholders' return may result in producing riskier and/or more volatile earnings streams for shareholders. This increased risk can reduce the extent to which such efforts actually increase shareholder value.

But he then goes on to make a more important point that the various ratios used by Ferrari in his breakdown of total shareholder return are unlikely to be independent. As a result, making changes to one such ratio could cause other ratios to rise or fall, making the effect on total shareholder return more ambiguous. An obvious example is that efforts to increase the premium to surplus ratio, by writing more business, could require writing less profitable business with lower U/P ratios. However, this can be avoided by incorporating strict underwriting guidelines.

But he gives three specific examples of more powerful relationships that exist when overall risk is left unchanged:

- Increases in the P/S ratio will likely lead to lower investment return on assets, I/A, for two reasons. First, as premiums increase more of the firm's assets will be tied up in non-marketable securities such as agents' balances. Second, in order to keep the overall riskiness of the company constant, taking on more insurance risk will require a shift to a more conservative investment strategy.

- P/S ratios will move in the same direction as the underwriting profit, U/P, since the more profitable the business the more insurance risk you can assume without increasing the risk of insolvency, for instance.

- Underwriting profits and investment return on assets will move in the same direction, as described earlier — higher underwriting profits allow for more risk-taking on the investment side, which is usually associated with higher average returns from marketable securities.

Practice Questions

Question 1. Write the formula for an insurers' total return on surplus, the total shareholder return, as a function of its insurance leverage and its insurance exposure.

Solution. First, we need to ensure we know the terminology that Ferrari used. Insurance leverage refers to the ratio of reserves to surplus, R/S; insurance exposure, which is another form of leverage, is the ratio of premium to surplus, P/S.

Second, total shareholder return is simply a function of its investment income, its underwriting income and its surplus, or:

$$\frac{T}{S} = \frac{I+U}{S}$$

A series of algebraic substitutions leads to:

$$\frac{T}{S} = \frac{I}{A}\left(1 + \frac{R}{S}\right) + \frac{U}{P}\frac{P}{S}$$

Question 2. An insurer writes \$75 million in premium at a combined ratio of 94%. If their invested assets earn an 8% return, their reserves are equal to \$150 million and their surplus is \$50 million, what is the total shareholder return?

Solution. We can use Ferrari's formula in terms of the reserve to surplus and premium to surplus ratios:

$$\frac{T}{S} = \frac{I}{A}\left(1 + \frac{R}{S}\right) + \frac{U}{P}\frac{P}{S}$$

To get the various values for each variable we notice that we were given $I/A = 8\%$ and we can easily get $U/P = 1 - \text{Combined Ratio} = 6\%$. Now we just need the leverage variables,

$R/S = 150/50 = 3$ and $P/S = 75/50 = 1.5$. Using the formula above:

$$\frac{T}{S} = 8\%(1+3) + 6\%(1.5) = 41\%$$

Of course, we could have also just calculated the intermediate values as $A = R + S = 150 + 50 = 200$, $U = 6\% * (75) = 4.5$ and $I = 200(8\%) = 16$. From this, $T = I + U = 20.5$ so the total return is $T/S = 20.5/50 = 41\%$.

Question 3. Redo the previous question assuming the combined ratio is 104%.

Solution. Here we have an underwriting return on premium of −4%, which is 8 percentage points lower than before. However, due to the leverage that comes from writing at a 1.5 premium to surplus ratio, plugging in to the same formula as before leads to a 26% shareholder return, which is 15 percentage points lower than before.

Question 4. Ferrari notes that in some sense an insurer's reserves reflect borrowed funds (borrowed from the policyholder) that can be invested. What does he say the "cost" of this borrowing is and how does it relate to the typical cost of funds for non-insurers who borrow from their debt investors? What is the cost of funds for the insurer in the previous questions?

Solution. The cost of these borrowed funds is the underwriting loss, which relative to premium was 4% in the previous question. Relative to reserves, this cost was 2%. This 2% cost of funds differs from the cost of actual borrowing because in the latter case the interest paid is usually fixed, whereas in this case the cost is really just an expected value. Actual costs could be much higher or lower.

We can easily recalculate the total return on surplus from above using Ferrari's second formula which reflects the investment income on reserves and the "cost" of borrowing these funds. We just need to be careful to use the underwriting profit relative to reserves, which is $U/R = -2\%$ here.

$$\begin{aligned}\frac{T}{S} &= \frac{I}{A} + \frac{R}{S}\left(\frac{I}{A} + \frac{U}{R}\right)\\ &= 8\% + (3.0)\,(8\% + (-2\%))\\ &= 8\% + 18\%\\ &= 26\%\end{aligned}$$

Question 5. Ferrari's formulas make it seem as though we could make the shareholders better off simply by increasing the reserve leverage, so long as the absolute value of I/A is greater than the absolute value of U/R. Is this correct?

Solution. It is true that increasing the reserve leverage leads to a higher expected total return for shareholders, but it also makes the shareholder returns more volatile. Increasing expected returns will lead to higher expected cash flows but if these are riskier they may be discounted using higher discount rates and may not lead to a higher actual value for shareholders.

Question 6. Some have observed that there are at times capacity problems in the insurance industry, with too little insurance being provided to society. This suggests that some people think the industry reserve to surplus ratio should be higher than it is. What three explanations does Ferrari give for this perception.

Solution. One explanation is that the reserve to surplus ratio that is optimal for the shareholders (and therefore targeted by company management) may simply be lower than the ratio that best serves society as a whole. This can give the perception of inadequate leverage even when the industry is operating at its optimal leverage.

Another is that the higher the riskiness of the investment portfolio the lower is the optimal reserve to surplus ratio. If the industry operates with too much investment risk, it will lead to a lower optimal reserve to surplus ratio and hence the perception of a capacity problem.

Finally, perhaps the comments are correct and the industry is indeed under-leveraged, but not because it isn't writing enough insurance but because it is writing it with too large of a capital base.

Question 7. In Balcareks review of the Ferrari paper he made the point that it might be more difficult than it appears to use Ferrari's breakdown of total shareholder return into its component parts to maximize shareholders' equity values and obtain an optimal level of insurance leverage. What three examples does he give of two ratios within Ferrari's formulas being dependent on each other?

Solution. The three examples given are:

- Increases in the P/S ratio will likely lead to lower investment return on assets, I/A, for two reasons. First, as premiums increase more of the firm's assets will be tied up in non-marketable securities such as agents' balances. Second, in order to keep the overall riskiness of the company constant, taking on more insurance risk will require a shift to a more conservative investment strategy.

- P/S ratios will move in the same direction as the underwriting profit, U/P, since the more profitable the business the more insurance risk you can assume without increasing the risk of insolvency, for instance.

- Underwriting profits and investment return on assets will move in the same direction, as described earlier — higher underwriting profits allow for more risk-taking on the investment side, which is usually associated with higher average returns from marketable securities.

Roth: Analysis of Rate of Return without Using Leverage Ratios

Introduction

In this reading, the following key points are made:

- It is not theoretically possible to allocate surplus by line of business and, as a result, no premium to surplus (leverage ratios) exist that can be applied to all insurers.

- The proper way to measure income is to reflect *all* sources of income, thus capturing the full change in surplus during the year (excluding the increases in surplus due to additional paid in capital and the decreases due to the payment of dividends).

- The required return for insurers, both stock and mutual, is the amount that will allow them to continue to provide the insurance capacity needed for society, which includes a provision for each of the following — inflation, changes in the valuation of the existing liabilities and changes in demand for insurance.

- Stock insurers require an additional component of return, which is the dividend that will ensure that their investors earn a reasonable overall rate of return on market value and that additional capital can be attracted as needed.

Allocation of Surplus to Line and State

Some argue that profitability for insurers should reflect profits relative to a base such as surplus. But whether imputed to lines of business and states through the use of leverage ratios (premium to surplus ratios) or determined by an allocation of the insurer's actual surplus, this process has significant problems.

The appropriate amount of aggregate surplus is unique to each insurer based on all of its risks, so leverage ratios can vary considerably by insurer. And since all of the surplus an insurer stands behind all of its risks, it cannot be allocated to line or state in a realistic fashion.

In addition, an insurer's actual surplus at any point in time will vary considerably from the theoretically appropriate amount of surplus needed for its risks.

As a consequence of not being able to meaningfully allocate surplus we also can't allocate investment income on surplus, which makes it impossible to measure total profitability by line or geography.

Measuring Income

Historically insurance rate regulation excluded investment income from measures of insurer profits. When interest rates rose considerably in the 1980's, this became difficult to justify. But even then, recommendations from an NAIC study continued to exclude certain sources of income including policy fees, unrealized capital gains (and in some cases even realized capital gains).

This doesn't make sense from a purely economic perspective — all income should be included, which means that it can be measured as the change in surplus (dS), excluding surplus changes that result from flows between the insurer and its shareholders (dividends and paid in capital). The rate of return then, when measured against surplus, is simply dS/S.

Table 1 shows Roth's calculations using the reconciliation items shown in the statutory income statement (Page 4 of the current annual statement blank). The entries shown in Table 1 are a bit out of date — there are now more entries in the reconciliation and the descriptions differ in some cases. Nevertheless, it is often confusing how the "changes" are calculated because in the statutory statement sometimes the change shows the increase and sometimes it shows the decrease.

TABLE 1. Calculation of Total Economic Income (dS)

dS =	Net Underwriting Gain or Loss
+	Net Investment Income
+	Net Realized Capital Gains or Losses
+	Other Income
-	Dividends to Policyholders
-	Federal Taxes
+	(Change in) Net Unrealized Capital Gains or Losses
+	Change in Non-Admitted Assets
+	Change in Liability For Reinsurance
+	Change in Foreign Exchange
+	Change in Excess Statutory Reserves
+	Other Write-In Items
=	Total Economic Income

Note the following clarifications:

- Change in Net Unrealized Capital Gains or Losses

  Increases are not included in the measurement of net income but they do appear in the valuation of the assets, so the increase has to be added to surplus to reconcile the balance sheet.

- Change in Non-Admitted Assets

  Consider what happens when you use cash to buy a non-admitted asset such as furniture. Your admitted assets decline but there is no income statement impact, and so the increase in the non-admitted "asset", which doesn't appear on the balance sheet, has to be subtracted from the surplus account to reconcile the balance sheet.

- Change in Liability For Reinsurance (Schedule F Penalty)

Increases in this provision do not impact the measurement of net income, but they are added to a "liability" entry on the balance sheet and are removed from the surplus account.

Example: We earn $100 of premium, incur $80 of gross loss and cede $30 to an unauthorized reinsurer. Net income will show $50 of income.

However, when calculating the balance sheet entries, the asset rise by $100 and the liabilities rise by $80 since we can't recognize the $30 cession (it will appear as $50 net loss reserve and $30 provision for reinsurance on the balance sheet). This causes surplus to rise by only $20. In order to reconcile the $50 of net income with the actual change in surplus we have to subtract the $30 increase in the provision for reinsurance.

- Change in Foreign Exchange

  This is just like the unrealized capital gains and reflects the foreign exchange impact on invested assets. Increases aren't included in income and so they have to be added to reconcile the change in surplus.

- Change in Excess Statutory Reserves

  This no longer exists.

GAAP vs. Statutory Accounting

If we measure income based on the change in net worth, does it matter whether we use statutory surplus or GAAP equity in this calculation? In practice, perhaps not if we can assume that GAAP equity and Statutory surplus differ by a fixed proportionality constant. In this case, dS/S will be roughly independent of the accounting standards used.

Fair and Reasonable Rate of Return

The US Supreme Court (in the Hope Natural Gas case) has ruled that a fair and reasonable rate of return for equity owners is one that is commensurate with returns in other enterprises with similar risks and that is sufficient to attract capital to the industry.

Using that logic, Roth argues that the best way to determine if returns for the industry are fair and reasonable is to observe whether the industry is attracting capital and new companies are being formed. The former criteria means that insurers are at least *expecting* to earn a fair and reasonable rate of return.

For instance, Roth calculates that from 1977-1990 the average rate of return for the industry was 14.1% and showed significant variability. It's hard to interpret the specific returns, but since the industry both paid dividends and attracted new capital each year during this period, we know that investors at least expected to earn a reasonable return, whatever that amount was from year to year.

How Much Capital is Needed?

If being able to attract sufficient capital to the industry is the best indication of fair and reasonable rates of return, how do we quantify how much capital is sufficient?

One approach, from an economist's perspective, may be to assume that the industry overall is in equilibrium with supply of insurance capacity sufficient to meet demand for insurance. If so, then the amount of surplus that will be needed *next year* is the amount that is sufficient to retain this equilibrium. This occurs when the change in surplus during the year is sufficient to provide for the following:

- expense and claim inflation
- increase in aggregate reserves
- increase in demand for insurance

The first two of these components affect the supply (capacity to assume risk) from the industry while the third affects the demand. Together, these represent the minimum amount needed to maintain equilibrium.

Quantifying the Required Surplus Change

To quantify the three elements above, Roth uses historical and forecasted relationships based on industry data.

To begin, expense and claim inflation can be estimated using historical rates of claim inflation, with whatever subjective adjustments seem appropriate to reflect the current economic environment.

To measure the increase in surplus required to support increases in demand for insurance, we can measure the real (inflation-adjusted) growth rate in premiums under the assumption that premium growth is highly correlated with risk-taking capacity growth. Such historical growth rates can be used, with whatever adjustments are appropriate to reflect the current economic environment, to capture the growth in demand. To calculate the growth in surplus required to meet this demand, if we assume that premium to surplus ratios are constant then the growth rate in surplus has to equal the growth rate in premium.

Finally, to measure the surplus required for the increase in aggregate reserves we can use the growth rate in industry loss reserves, also on an inflation-adjusted basis (since expected inflation has already been explicitly taken into account). But here, we need to be careful and note that observed reserve growth rates reflect the growth in both old and new business. Since we are explicitly including the surplus needed for the new business, we only want to include the reserve growth related to existing business each year. We can approximate this by subtracting the historical growth rate in premiums from the historical growth rate in reserves.

Stock Insurers

Notice that if surplus were to grow during the year by an amount equal to the three items above there would not be any ability to pay dividends to shareholders. As a result, we need to add the amount of dividends required by investors to these amounts to ensure that shareholders, the ultimate providers of the insurance capacity, earn a return satisfactory to them and sufficient to attract capital to the industry when needed (for instance, when despite the fact that the policies were priced with underwriting profit margins sufficient to grow surplus enough on an expected basis, actual results were worse than expected and so new capital must be provided to retain industry equilibrium).

Mutual Insurers

Notice that the last component identified, the dividends to stockholders, does not apply to mutual insurers. They do not need to pay “shareholders” nor do they have a need or ability to attract capital, so only the first three components listed above apply to them.

Stockholder Returns and the Market-to-Book Ratio

Notice that even though we’ve been talking about returns to stockholders, we’ve actually been using returns relative to surplus, or book value. This is odd because what the stockholder has invested, the value of his holdings that he could get if he sold them, is the market value. In reality, what stockholders care about is the return they earn on the market value of their holdings.

At the time the Roth paper was written, typical P&C companies’ ratio of market value to book value was around 2.0. Today, it is much closer to 1.0 (or lower). Nonetheless, when assessing returns to stockholders the rate of return should be converted to a market value basis[1].

More importantly, Roth points out that in a competitive capital market, the market value of stock insurance companies should constantly adjust so that, by definition, its shareholders are earning a fair and reasonable rate of return on market value. If the return on surplus is not fair and reasonable, then the market-to-book ratio will adjust accordingly, capital will flow in, dividends will be paid out, etc. This argues that at any point in time the rate of return on surplus must be the competitive equilibrium rate of return. By extension, changes in the market-to-book ratios should provide an indication of changes in rate adequacy.

Practice Questions

Question 1. Roth says that the proper measure of return for insurers should be analyzed using the “general principles of economics” and “actuarial analysis of the structure and trends in the insurance industry”. What does he mean by these two terms?

[1] Notice too that actual stockholder returns are impacted by *changes* in the market to book ratios.

Solution. Roth believes that the proper amount of income for insurers is the amount that will allow them to continue to provide the amount of insurance capacity demanded by society, which is a "general principle of economics" that supply should equal demand.

To assess the exact amount required it is necessary to quantify the effects of inflation, changes in demand and changes in the valuation of existing liabilities will impact the total surplus required. Such analysis will require "actuarial analysis of the structure and trends in the insurance industry".

Question 2. Given the following information for the P&C insurance industry, calculate the rate of return using Roth's methodology. All amounts are in billions of dollars and both dividends and paid-in surplus were paid during the year.

Beginning Surplus	271.5
Ending Surplus	301.5
Dividends Paid	25.0
Paid-in Surplus	15.0

Solution. Roth recommends calculating the profit for the industry inclusive of all sources of profit, so he looks at the total change in surplus during the year. In this case, that equals 30.0.

However, when the surplus is measured at the end of the year, it reflects the decrease from dividends paid and the increase from paid-in capital. Since we only want the income portion of the change in surplus, we have to reverse those adjustments by adding in the dividends and subtracting the paid-in capital. As shown below, those calculations result in a change in surplus of 40.0.

The rate of return then is calculated relative to the beginning surplus, so the return is 40.0/271.5 = 14.7%.

Beginning Surplus	271.5
Ending Surplus	301.5
Increase in Surplus	30.0
Add Back Dividends Paid	25.0
Subtract Paid-in Surplus	-15.0
Change in Surplus (dS)	40.0
Return on Surplus	14.7%

Question 3. Suppose we wanted to calculate the rate of return from the previous question using GAAP equity rather than Statutory surplus. If the ratio of GAAP equity to Statutory surplus is generally equal to 1.20, how would you approximate this return

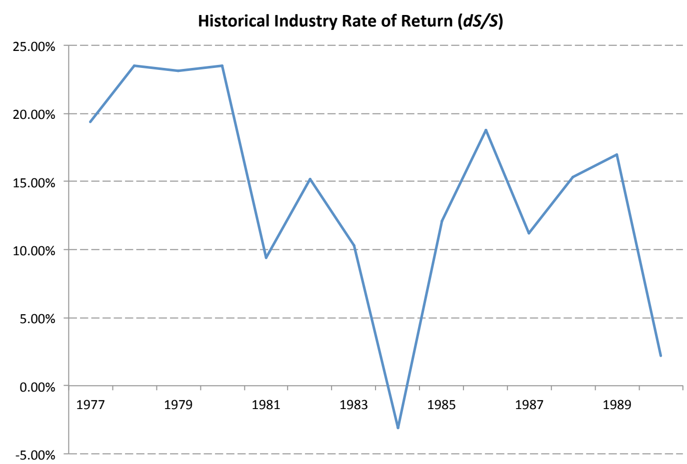
Solution. Here, we simply adjust the beginning and ending surplus to be on an approximate GAAP basis by multiplying by 1.20 and then continue with the same calculations as before.

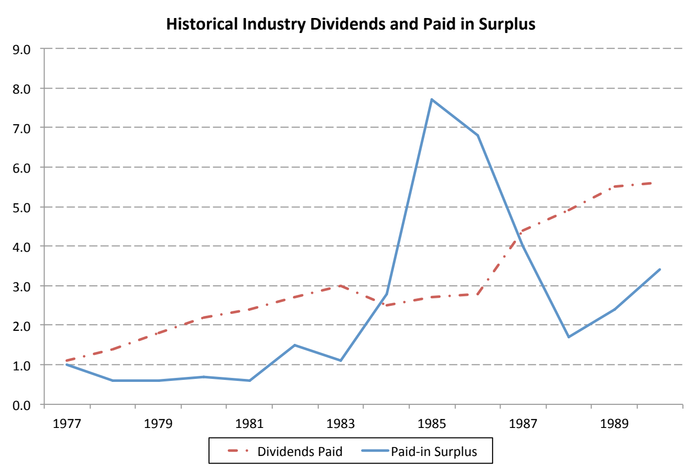
Neither the dividends nor the paid-in surplus amount are adjusted, since those are dollar amounts independent of the accounting basis.

The resulting return is 14.1%, which is not too different from the return calculated on a statutory basis.

Beginning Surplus	325.8
Ending Surplus	361.8
Increase in Surplus	36.0
Add Back Dividends Paid	25.0
Subtract Paid-in Surplus	-15.0
Change in Surplus (dS)	46.0
Return on Surplus	14.1%

Question 4. The two graphs below show the annual rate of return on surplus for the insurance industry as well as the dollar amounts (in billions) of dividends and paid-in surplus each year. What aspects of these graphs suggest that the industry returns were fair and reasonable?





Solution. Although actual industry returns seemed to be high (over 10% in most years), we know that insurance profits are hard to measure accurately and likely to be volatile from year to year. It is not possible to infer from the historical return data whether investors expected to earn fair and reasonable returns each year.

However, the dividend and paid-in surplus data show that these amounts were positive every year, even when actual returns were low or negative. It is that data which leads us to believe that investors expected to earn fair and reasonable returns each year. The willingness for investors to provide more capital to insurers is the best indication of this.

Question 5. You are given the following facts about the insurance industry's historical and recent performance. You are trying to assess the profitability of stock insurers using Roth's approach. Determine both the (*a priori*) required return (as of the beginning of the current year) and the actual return. Assume that average historical growth rates provide reasonable estimates of prospective growth rates.

Historical Industry Performance (Stocks and Mutuals)		
Claim and Expense Inflation Rate	3.0%	
Written Premium Growth Rate (inflation-adjusted)	2.5%	
Reserve Growth Rate (inflation-adjusted)	6.0%	
Premium to Surplus Ratio	1.5	
Reserve to Surplus Ratio	2.0	
Current Year Data (in $ billions)	Dollars	% of Surplus
Beginning Surplus	450.00	
Ending Surplus	485.00	
Change in Surplus	35.00	7.78%
Stockholder Dividends Paid	10.80	2.40%
Paid-in Surplus	6.75	1.50%

Solution. Let's begin with the calculation that should be the easiest, which is the required surplus change based on industry trends. We were told to use historical average figures for the three components we need — inflation, growth in demand and growth in reserves. The first two are easy as they were in the question (along with a bunch of extraneous information — sorry!).

For the inflation we use the 3% historical amount and for the growth in demand we use the growth in inflation-adjusted premiums, which is 2.5%. For the reserve growth, we can see that the historical reserve growth was 6%, but that reflects both old and new business. To get the reserve growth from existing business only, we subtract the premium growth rate to get 3.5%. The net required surplus growth rate is 9%.

Notice that we don't use the leverage ratios provided since when premiums or reserves grow at a given rate, the surplus just changes proportionately and so the growth rate in surplus from these two causes is the same as the respective premium and reserve growth rates.

The 9% just calculated is the required growth rate in surplus, not the required return on surplus. To get the required return we need to also factor in the need to pay dividends and include anticipated paid-in capital. This is tricky because we weren't given any expected values as of the start of the year. What Roth did (in his Table 5) is assumed that the actual for the year was what was expected as of the start of the year.

This gives us a total required return of 9.9%, as shown below:

Required Surplus Change	9.00%
Add Expected Dividends Paid	2.40%
Subtract Expected Paid-in Surplus	-1.50%
Required Change in Surplus (dS/S)	9.90%

For the actual return, I will first show it using the approach we used in an earlier question, which will be the least confusing.

From the figures given in the table we have:

Actual Surplus Change	7.78%
Add Dividends Paid	2.40%
Subtract Paid-in Surplus	-1.50%
Actual Change in Surplus (dS/S)	8.68%

Comparing the required and actual returns, actual returns were 1.22% below the required return.

Notice that in Roth's Table 5 he showed this calculation a little bit differently, so it might be good to see what he did there. Basically, he backed into an intermediate value and then used that value in his final calculation.

We were told that the actual surplus grew by 7.78% and that the paid-in surplus was 1.5%. This means that without the paid-in surplus the surplus would have grown by $7.78\% - 1.5\% = 6.28\%$. This figure is what Roth calls "retained return on capital" in his Table 5. It is the portion of the return that does not reflect the fact that dividends have been paid as well. Then, adding in the dividends of 2.4%, we have the total actual return of 8.68%.

The following two calculations are similar to the ones shown in Roth's Table 5:

Retained Return on Capital	6.28%
Paid-in Surplus	1.50%
Actual Surplus Change	7.78%
Dividends Paid	2.40%
Retained Return on Capital	6.28%
Actual Return	8.68%

Question 6. If you were doing the same calculation of the required return as in the previous question, but this time for a mutual company, what would be the required return on surplus? What assumptions might you change?

Solution. The distinction for mutual companies is that they do not have to provide for dividends in their returns, in which case the required return on surplus is simply the required surplus change taking into account inflation, growth in demand and growth in reserves. Using the same figures as before, this would be 9%.

However, keep in mind that the mix of business for mutual and stock insurers is very different. The one change that Roth made was to reflect the fact that mutual companies write much more short-tailed personal lines business and so he adjusted the expected inflation rate accordingly. He didn't use stock-specific and mutual-specific historical premium and reserve growth estimates, but assuming you could get comfortable with the reliability of the data you could do that as well.

Question 7. What two reasons are given by Roth why regulators should not allocate surplus, and by extension the investment income on surplus, to lines of business or states?

Solution. The appropriate amount of aggregate surplus is unique to each insurer based on all of its risks, so leverage ratios can vary considerably by insurer. And since all of the surplus an insurer stands behind all of its risks, it cannot be allocated to line or state in a realistic fashion.

Question 8. You are given the following information from the financial statements of a stock P&C insurer which had starting surplus at the beginning of the year equal to $S_0 = 117,935$:

Net Underwriting Gain or Loss	(16,895)
Net Investment Income	31,207
Net Realized Capital Gains or Losses	4,649
Net Unrealized Capital Gains or Losses	8,035
Other Income	(1,228)
Change in Non-Admitted Assets	43
Change in Liability for Reinsurance (i.e. Schedule F Penalty)	(702)
Change in Foreign Exchange (on Invested Assets)	29
Change in Excess Statutory Reserves	195
Other Write-In Items	299
Federal Taxes	2,802
Dividends to Policyholders	2,713

Determine the statutory net income, the total economic income and the total return on surplus as calculated by Roth for this insurer.

Solution. First, statutory net income reflects UW income, investment income, realized capital gains, other income items, policyholder dividends (as an expense) and taxes. This would give statutory income of $12,218$.

Then to get the total change in surplus we subtract the federal taxes and dividends to policyholders and add all of the remaining items. Doing this gives total economic income of $dS = 20,117$.

The total return is then $20,117/117,935 = 17.06\%$.

A few things are worth noting:

- The excess statutory reserves don't exist anymore, but I left them in the question so that the question was consistent with the reading. Actual exam questions may or may not include them.

- The entries here use the same terminology as in the statutory statement but the meaning of the word "change" varies by entry. In some cases the change is the *increase* during the year and sometimes it is the *decrease* during the year. Just be careful and refer to the discussion in the notes for the details.

- In this question I showed the various amounts in a somewhat jumbled manner just to make you think about what to include. In the reading, the items appear more like the way they would appear on Page 4 of the Statutory Statement (in the section that reconciles the change in surplus during the year):

dS =	Net Underwriting Gain or Loss	(16,895)
+	Net Investment Income	31,207
+	Net Realized Capital Gains or Losses	4,649
+	Other Income	(1,228)
-	Dividends to Policyholders	2,713
-	Federal Taxes	2,802
+	Net Unrealized Capital Gains or Losses	8,035
+	Change in Non-Admitted Assets	43
+	Change in Liability For Reinsurance	(702)
+	Change in Foreign Exchange	29
+	Change in Excess Statutory Reserves	195
+	Other Write-In Items	299
=	Total Economic Income	20,117

The first six items comprise the statutory net income and the remaining represent adjustments to surplus shown in the statutory statement.

In addition, it is worth noting how the reconciliation is currently laid out, as there are different descriptions (e.g. the net unrealized capital gains are now referred to as the change in net unrealized capital gains) and additional items (e.g. change in surplus notes):

TABLE 2. Current Statutory Statement Layout

Line	Description
21	Surplus as regards policyholders, December 31 prior year
22	Net income
23	Net transfers (to) from Protected Cell accounts
24	Change in net unrealized capital gains (losses) less capital gains tax
25	Change in net unrealized foreign exchange capital gain (loss)
26	Change in net deferred income tax
27	Change in non-admitted assets
28	Change in provision for reinsurance
29	Change in surplus notes
30	Surplus (contributed to) withdrawn from protected cells
31	Cumulative effect of changes in accounting principles
32-34	Paid-in Capital
35	Dividends to stockholders
36	Change in treasury stock
37	Aggregate write-ins for gains and losses in surplus
38	Change in Surplus (Sum of Lines 22 to 37)
39	Surplus as regards policyholders, December 31 (Line 21 + Line 37)

Roth's definition of total economic income would essentially be Line 38 - Lines 32-36, because he would ignore the transactions between the company and the shareholders in Lines 32-36.

McClenahan: Insurance Profitability

Introduction

This paper looks at profitability from a regulated rate filing perspective and makes the case that, in that context, profit targets should be set relative to premium ("sales"). This has two very appealing practical benefits — it is easy to interpret and it does not require the use of allocated surplus to measure expected profitability.

Further, he argues that rather than rely on complex financial models that purport to determine appropriate rates of return on equity for shareholders, a less quantitative approach can be taken. Rates of return on surplus that are fair and reasonable will be those that lead to desirable market characteristics in terms of the size of the residual market, the degree of competition and the degree of product diversity and innovation.

Role of Investment Income

In a regulated rate environment, the author argues that policyholders should be given credit for the investment income expected to be earned on the *policyholder supplied funds*. We will define this more rigorously in the Robbin reading notes, but just note that this reflects the net premium flows held in reserves (unearned premium and loss reserves).

The investment income for this component should be on a risk-free rate basis to reflect the fact that any investment risk assumed in the investment process is borne by the shareholders (unless it causes insolvency, of course) and so earnings above the risk-free rate should not accrue to or impact policyholder rates.

Note that here, all investment income on surplus is also excluded.

Calculating Expected Profit

Profit in this case is measured, therefore, as premiums *less* losses and expenses *plus* investment income on policyholder supplied funds. The easiest way to calculate this though is to simply calculate present values of the premium, loss and expense cash flows on a risk-free present value basis. This will produce the present value profit.

One thing to note is that not all premium funds, even net of acquisition expenses, are invested in marketable securities that can earn (at least) the risk-free rate. Some investments are needed in physical assets and infrastructure, so in practice the premium should be adjusted to account for only the portion that can be invested in marketable securities.

Nonetheless, this simple calculation for regulatory rate purposes is appealing because it does not depend on the level of actual or assumed surplus, does not depend on actual or assumed investment results and does not depend on past underwriting experience.

Finally, keep in mind that this calculation of profit does not fully reflect the expected income for the insurer as it does not include expected investment returns above the risk-free rate on policyholder funds nor does it include investment income on surplus.

Rate of Return on Equity

So far, we focused on the dollar profit, on a present value basis and not any measure of a rate of return. This makes sense in a regulated rate environment because the role of the regulator is to assess the fairness of the profit load in the insurance rates. As such, insurers charging the same premiums should be treated the same, independent of their respective equity bases. Taking into account the equity base might be an important consideration for the investors, but it shouldn't impact the regulation of insurance rates.

Aside from that theoretical argument against using the rate of return on equity, there is a practical one as well. Since all of the insurer's surplus stands behind each of their policies, it is not possible to allocate portions of that surplus to individual lines of business or states. Besides the fact that such allocation is artificial, allocation of surplus can cause strange results from a rate regulation perspective.

Consider, for instance, two insurers writing the same business in the same state. Insurer *A* writes only in that state and is capitalized with \$1 million of surplus; insurer *B* writes nationally, has \$100 million of surplus in total and has allocated \$1 million to the state in question. If both *A* and *B* charge the same rates (on allocated surplus) no value will be given to the fact that *B* offers a better product because it is backed by more capital.

Rate of Return on Sales

It is sometimes argued that a convenient way to allocate surplus to lines and states is to apply benchmark premium to surplus ratios. Obviously this does nothing to address the practical issues with allocating surplus, it just turns it into a question of determining the appropriate benchmark premium to surplus ratio. Even more fundamentally, this is just an artificial way to regulate the rate of return on sales (premium) by multiplying the result by a constant ratio.

Why not just make things clearer and focus on the rate of return on premium to begin with? This has two benefits.

First, it provides useful information to the policyholders because there is no ambiguity in interpreting the figure. Telling someone that rates reflect a 5% markup on premium for profit is clear; telling them that rates reflect a 12.5% return on equity is not.

Second, implementing a return on premium benchmark is easy and doesn't introduce ambiguities for the regulator as a result of the surplus allocation issues.

Determining the Benchmark

Measuring historical insurance profitability is inherently difficult given the uncertainties associated with estimations of the loss components, even years after the business is written. So how is a regulator supposed to establish a benchmark return on sales rate? McClenahan recommends relying on observable market behavior to discern whether insurers perceive that prospective returns will be reasonable.

Because insurers are free to react to perceived rate inadequacies through changes in underwriting standards and volume, observing the composition of the residual market, the number of insurers in the voluntary market and the degree of product diversity and innovation provides insight into whether insurers perceive regulated rates to be reasonable. Regulators can therefore establish return on premium standards as the rates that produce the desired market characteristics along these dimensions.

Practice Questions

Question 1. What are two advantages of using return on premium as opposed to return on equity mentioned by McClenahan for the purposes of regulated insurance rate setting.

Solution. Return on premium has two practical benefits — it is easy to interpret and it does not require the use of allocated surplus to measure expected profitability.

Question 2. For the purposes of setting regulated insurance rates, how does McClenahan say investment income should be taken into account?

Solution. He argues that only the investment income, at the risk-free rate of return, on policyholder supplied funds should be included. Investment income on surplus, as well as returns over and above the risk free rate on policyholder supplied funds, belong to the shareholders and should not be used to offset the profit provisions in rates.

Question 3. Other authors often argue that insurance rates should be set so that the insurers' shareholders can expect to earn a return commensurate with the risk they assume. Establishing this target rate of return often requires reference to returns available on other investment opportunities or theoretical models such as CAPM.

How does this compare to McClenahan's recommendation for how to set regulated rate targets which can be said to be reasonable, not excessive and not inadequate. Comment on how these different recommendations can be reconciled.

Solution. McClenahan argues that regulators should assess the reasonableness of rates relative to desired market characteristics such as the size of the residual market, the degree of competition and the degree of product diversity and innovation. That is, as insurers compete in the market they will respond through underwriting standards and premium volume to their own assessments of the fairness of rates (from their point of view). Insurers unwilling to write a particular class of risks is a clear indication that, in the eyes of the insurers, rates are inadequate.

Notice that McClenahan's approach ultimately depends on the behavior of the insurers themselves. If the insurers evaluate premiums using return on equity benchmarks, then regulating rates based on simpler return on sales measures may actually not be very different from using return on equity benchmarks. But it will be simpler for regulators.

Note also that when reading the papers on pricing insurance from the investors' perspective, we are typically quantifying *minimum* risk loads sufficient to attract capital and just satisfy shareholders. The implicit recommendation is to charge as much as the policyholder will pay, but no less than the premium determined in this manner. On the other hand, McClenahan's recommendations relate to how regulators should establish *maximum* rates allowed.

Robbin: Underwriting Profit Provision

Introduction

The Robbin paper outlines seven methods for determining an *underwriting profit margin* (U) for the pricing of a single line of business on a prospective basis. Each of his methods are surprisingly simple and intuitive. The challenge in this paper is to keep track of the notation and variable names, and in particular the recommended sources for the needed inputs.

A more subtle challenge is to pick up insights into how and why these methods differ from each other, which I will attempt to focus on a bit to better prepare you for exam questions.

In the notes here, I want to emphasize the simplicity of the methods as much as possible. To do so, I will first show the formulas without the use of any particular variable names and then show the formulas using Robbin's notation. I will generally avoid commentary about the subtleties of using actual data to calculate the inputs and instead will use numerical problems to demonstrate these calculations.

Overview of Methods

The seven different methods for determining the U can be categorized into three broad approaches. Without using any formulas or notation, it will be helpful to consider a general overview of these three approaches. Subsequent sections will explore the seven specific methods, and their nuances, separately.

Investment Income Offset

These methods start with the traditional underwriting profit margin (U^0), which had historically been a flat 5% loading for all lines (2.5% for workers' compensation), and adjust it to take into account investment income that is earned on essentially the funds provided up-front by the policyholder.

For example, policyholders pay the full premium up-front but it is earned over time. The unearned premium balance can be thought of as belonging to the policyholder until it is earned and therefore the profit margin should be *reduced* by the investment income that will be earned on those funds.

Similarly, a substantial portion of the premium is intended to pay expected claims. While these funds sit within loss reserve accounts of the insurer, they can again be thought of as belonging still to the policyholder. Investment income earned on these funds should serve to reduce the premium charged and hence should serve as an offset to the U.

Two specific methods using this approach are presented. In one, the offset is comprised of the expected investment income rate (after-tax) and the total amount of policyholder supplied funds. In the other, the offset is comprised of the risk free rate and the income that could be generated from the loss reserve funds *relative to* a benchmark line of business. That is, the standard $U^0 = 5\%$ is deemed to be fair for a benchmark line of business and lines that have faster (slower) loss reserve payment patterns will have higher (lower) values for U.

Target Total Return Methods

These methods take the perspective of the investor and seek to ensure that the pricing will lead to a target expected total shareholder return. Because the total return is the focus, these methods take into account investment income earned on all assets, including shareholder capital (surplus).

Three specific methods are discussed within this category. One method looks at a simple definition of the return on equity (ROE) as would be measured by accountants and solves for the U such that the expected ROE equals the target ROE. The other two methods capture the actual cash flows *from the investors' perspective*, tracking the contribution of equity capital (as if the line of business or policy were the only one written by the company) at the outset and the flow of profits and capital, including investment income, back to the investors over time.

These two equity flow methods differ only with respect to the mechanics of the calculated rate, where one is an internal rate of return (IRR) and the other a ratio akin to the ratio of profits to invested capital. In fact, under certain conditions the two result in identical underwriting margins.

Cash Flows Methods

The last category differs from the others in that rather than use accounting measures of income it relies on the actual premium, loss, expense and tax cash flows.

In one method, it is argued that to attract equity capital to an insurance company in support of writing a policy, the present value of the investors' equity flows must be equal to (or less than) the present value of the after-tax underwriting cash flows. In this approach, Robbin discounts the equity flows at the cost of capital, or the target rate of return, and the underwriting cash flows at the investment yield.

In the other method, the Capital Asset Pricing Model (CAPM) is used to determine a risk-adjusted discount rate for the liability, expense and tax cash flows and then the U is set so that the present value of the premium is equal to the risk-adjusted present value of these cash flows.

Seven Specific Methods — Comment on Presentation

With the above overview, hopefully it is clear that each of the methods discussed are simple and intuitive. Now all that remains is to show some formulas. Well, that's not entirely true. It would be helpful to also comment on the practical issues involved, the inconsistencies encountered, etc. I will reserve most of those comments for the next session though, since students find this distracting commentary the source of so much confusion in the Robbin paper.

In what follows, I will retain the categorical outline presented before and develop the formulas needed for each of the seven methods. The numerical examples included here will demonstrate the ease with which these approaches can be implemented *when given the inputs needed.* The subtle aspects of calculating each of the inputs are best left to numerical problems at the end of the reading, which is where most of it will appear.

Finally, be sure to keep the relationship among the methods — where they fit into the three categories described above — in perspective. I will make this easy by presenting similar methods sequentially, but please refer to Table 1 as needed.

TABLE 1. Summary of Seven Methods

Approach	Method
Investment Income Offset Methods	CY Investment Income Offset
	Present Value Offset
Target Total Return Methods	CY ROE
	IRR on Equity Flows
	PV of Income to PV of Equity Ratio
Cash Flow Methods	PV Cash Flow Return
	Risk-Adjusted Discounted Cash Flow

Calendar Year Investment Income Offset

Here, we simply want to start with the traditional $U^0 = 5\%$ and reduce it by the investment income from the policyholder supplied funds. As noted before, the policyholder supplied funds reflect both of the following:

- the unearned premium reserve, which is net of the associated acquisition expenses already paid by the insurer and net of any premium receivables (unpaid premiums)
- loss and LAE reserves

Further, the investment income rate reflects the total, after-tax investment portfolio rate.

A point to keep in mind, which applies to all of the methods that will be presented, is the fact that the underwriting profit is being calculated relative to premium (i.e. as a ratio) and not in dollars. Therefore, when we refer to offsets based on reserves we need to use reserves relative to premium as well. This won't always be obvious in the notation used by Robbin.

The Key Formula

Denoting the traditional U as U^0, we have:

$$\begin{aligned} U &= \text{Traditional U} - \text{After-Tax Inv Income on Policyholder Supplied Funds} \\ &= U^0 - i_{AT}[\text{Policyholder Supplied Funds}] \\ &= U^0 - i_{AT}(PHSF) \end{aligned}$$

where,

$$\begin{aligned} PHSF &= \text{Unearned Premium Balance} + \text{Loss and LAE Reserves} \\ &= [\text{UEPR Net of Prepaid Expenses} - \text{Premium Receivables}] \\ &\quad + [\text{Loss Ratio} * (\text{Loss Reserves/Incurred Loss})] \end{aligned}$$

As a reminder, the policyholder supplied funds are calculated relative to premium so that both the U^0 and the offset calculations are in terms of premium ratios. The numerical problems at the end of the reading will explore the authors' suggestions for calculating the required inputs from actual data, including when to smooth results over multiple calendar years, etc.

Numerical Example

Suppose that over the past three years an insurer has had an average of $160,000$ in calendar year earned premium, $50,000$ in unearned premiums, an average of $28,000$ of premiums receivable on its balance sheet and a pre-paid acquisition expense ratio of 18%. Also, over the same three year period, the average ratio of loss reserves to losses incurred has been 1.20. Its current after-tax investment portfolio yield is 6.68% and its target loss ratio used in pricing is 60%. Determine the policyholder supplied funds relative to premium and the underwriting profit provision using the CY Investment Income Offset method and a 5% traditional underwriting profit margin.

We start by calculating the policyholder supplied funds from unearned premiums. We use the average dollar value of the unearned premium net of acquisition expenses and the average dollar value of the premium receivables to get:

$$\begin{aligned} \text{Unearned Premium Balance in Dollars} &= UEPR(1 - PPACQ) - RECV \\ &= (50,000)(1 - 18\%) - 28,000 \\ &= 13,000 \end{aligned}$$

Writing this relative to the average premium we have a balance of policyholder supplied funds from the unearned premium of $13,000/160,000 = 8.13\%$.

For the loss reserves, we see that the ratio of reserves to incurred losses is 1.20. WIth a 60% permissible loss ratio, these reserves relative to premium would be:

$$\begin{aligned}\text{Reserve Balance Relative to Premium} &= PLR\left(\frac{LRES}{INC}\right)\\ &= .60(1.20) = 72\%\end{aligned}$$

From this, we see that starting with $U^0 = 5\%$ we get an underwriting profit margin by subtracting the expected investment income yield on the policyholder supplied funds of:

$$\begin{aligned}U &= U^0 - i_{AT}(PHSF)\\ &= 5\% - 6.68\%(8.13\% + 72\%)\\ &= -0.35\%\end{aligned}$$

Permissible Loss Ratio Calculation

Notice that the previous example used the expected (permissible) loss ratio to calculate the U, but the U actually determines the expected loss ratio. This suggests that the solution must actually be done iteratively.

Advantages and Disadvantages

This approach is simple to implement using readily available data. The results are relatively stable, especially when capital gains are excluded from the investment income calculation, unless there is rapid growth or decline in volumes that distort some of the ratios used.

Present Value Offset

For this method we assume that the traditional underwriting profit margin, $U^0 = 5\%$ is appropriate for a *typical* line of business. We then develop an offset factor that reflects whether a given line of business has a faster or slower loss payment pattern than this, so that lines for which more investment income can be generated from the loss reserves will have lower target underwriting profit margins.

The Key Formula

If we denote the annual loss payment pattern for the reference line and the line under review as X^0 and X, then we can write present value difference in the investment income under each payment pattern as $L[PV(X^0)] - L[PV(X)]$. This quantity will be positive if the line under review pays out slower (earns more investment income on the reserve balances) than the reference line of business, in which case it should be subtracted from the premium charged.

Before showing the formula for the underwriting profit margin shown in the reading, let's now go back to first principles and write the premium that we would want to charge as being a simple sum of the discounted losses, expenses, the traditional profit margin in dollars *plus* the investment income that can be earned on the loss reserves under the reference payment pattern:

$$\begin{aligned} P &= L[PV(X)] + EXP + U^0(P) + L[1 - PV(X^0)] \\ &= \frac{L + EXP - L[PV(X^0) - PV(X)]}{1 - U^0} \end{aligned}$$

In that equation, the term $L[1 - PV(X^0)]$ represents the difference (in today's dollars) between the undiscounted loss and the discounted loss, which reflects the amount of investment income on the losses. The rationale for adding this into the equation above is that we are also assuming that the traditional underwriting profit margin, $U^0(P)$, already includes an "offset" for this amount.

In any case, take that logical formula for the premium and rewrite it as the underwriting profit margin in dollars:

$$\begin{aligned} U_{\$} &= P - L - EXP \\ &= [L[PV(X)] + EXP + U^0(P) + L[1 - PV(X^0)]] - L - EXP \\ &= U^0(P) - L[PV(X^0) - PV(X)] \end{aligned}$$

Finally, we want to show this as a ratio of premium, so simply dividing by P gives us:

$$U = U^0 - (PLR)[PV(X^0) - PV(X)]$$

This is Robbin's formula for what he calls the present value offset.

However, notice that in the last step I divided by the premium. If you had to solve for the premium, using a formula for U as a function of the PLR doesn't seem too helpful. You can use it, but you wind up having to write the U in terms of P and do some extra algebra.

I believe the previous formulas are more useful and capture the point of the method, which is to adjust a traditional underwriting margin by the difference in the present value factors for the reference line and the reviewed line.

Numerical Example

Suppose the undiscounted expected losses for a line of business are \$70 and there are no expenses. You want to use the PV Offset method to determine the underwriting profit provision given a traditional underwriting profit provision of $U^0 = 8\%$. Assume the reference line of business has all losses paid at the end of year 1, the line of business under review has all losses paid at the end of year 2 and the discount rate for losses is 5%.

To do this calculation, we need to know the difference in the present value factors for the two lines of business:

$$PV(X^0) = \frac{100\%}{1.05^1} = 0.9524$$

$$PV(X) = \frac{100\%}{1.05^2} = 0.9070$$

$$\Rightarrow [PV(X^0) - PV(X)] = 0.0454$$

Then, the premium under this method is simply:

$$\begin{aligned} P &= L[PV(X)] + EXP + U^0(P) + L[1 - PV(X^0)] \\ &= L + EXP + U^0(P) - L[PV(X^0) - PV(X)] \\ &= \frac{L + EXP - L[PV(X^0) - PV(X)]}{1 - U^0} \\ &= \frac{70 - 70 * .0454}{1 - .08} \\ &= 72.64 \end{aligned}$$

Now that we know what premium to charge, we can determine the proportional underwriting profit margin simply as:

$$\begin{aligned} U &= 1 - \text{Combined Ratio} \\ &= 1 - 70/72.64 \\ &= 1 - 96.37\% \\ &= 3.63\% \end{aligned}$$

Note that this isn't the formula Robbin showed, so exam questions might be presented differently. For instance, if I had told you the expected loss ratio (what Robbin calls the permissible loss ratio, PLR) then you could have found U as:

$$\begin{aligned} U &= U^0 - (PLR)[PV(X^0) - PV(X)] \\ &= 8\% - (96.37\%)[.0454] \\ &= 3.63\% \end{aligned}$$

Discount Rate

When discounting the losses in this method, current yields on new investments (the *new money rate*) is more theoretically sound. But Robbin argues that in practice existing portfolio yields will tend to be more stable and might be easier to justify to regulators.

Taxes

To partially account for the effect of taxes, an after-tax rate could be used for discounting.

Advantages

This method uses a simple method to reflect an investment income offset and does so without relying on historical ratios of balance sheet items that could be distorted by changes in volume. It also avoids having to select a target return or allocate surplus to a line of business, as other methods we will show do.

Calendar Year ROE

This method simply calculates a calendar year ROE and sets the premium so that a target ROE is achieved (prospectively of course, so on an expected basis).

Key Formula

The ROE is simply the expected after-tax income over equity. Since the expected income is premium less losses, expenses and taxes, this can be written simply as:

$$
\begin{aligned}
ROE &= \frac{\text{UPP in Dollars} + \text{Investment Income} - \text{Taxes}}{\text{GAAP Equity}} \\
&= \frac{U \cdot P + II - FIT}{EQ}
\end{aligned}
$$

Then we solve for U so that the expected ROE equals the target ROE.

It should be noted that rather than separate the total taxes as shown above, we could just use an after-tax underwriting profit and an after-tax investment income rate, but the logic is the same.

Numerical Example

Working through a numerical example could either be trivial or difficult, depending on what information is given to you. The formula shows how easy the calculation would be if you were given the target ROE and all of the inputs. But on the exam you may be given more basic values and you may have to calculate the key inputs in the above equation.

For instance, suppose you are told that only the following:

- The company writes 103.35 in premium.
- The expected underwriting income is −2.41% of premium.
- They pay a 34% tax rate on underwriting income (and get a refund at that rate for underwriting losses).

- The policyholder supplied funds (from both unearned premium and loss reserves) 80.13% of premium.
- The premium to surplus ratio is 3.0.
- The expected portfolio yield on an after-tax basis is 6.68%.
- The ratio of GAAP equity to Statutory Surplus is 1.2.

From this, if we wanted to calculate the CY ROE we would not be able to directly use the formula shown above because we don't know the dollar amount of the investment income after tax. We can get this by noting that we earn investment income on both the policyholder supplied funds and the surplus, so:

$$\text{After-Tax Inv Income} = 6.68\%[(80.13\%)(103.35) + 103.35/3.0] = 7.83$$

With this, we can calculate the after-tax income as:

$$\begin{aligned} INC &= (1 - t_u)U \cdot P + II_{AT} \\ &= (1 - .34)(-2.41\%)(103.35) + 7.83 \\ &= 6.19 \end{aligned}$$

Now, to get the CY ROE, we also need the GAAP Equity. We are told that the ratio of GAAP Equity to Statutory Surplus is 1.20, so we can calculate $EQ = 1.20(103.35/3.0) = 41.34$. This gives us our result:

$$\text{CY ROE} = \frac{6.19}{41.34} = 15\%$$

Suppose that our target ROE were 20%. How would we set the underwriting profit margin? We would just have to solve for the profit margin in the premium so that the ratio as calculated above equalled 20%.

I don't think it is necessary to know this formula. But if you want to be sure you don't have to do the algebra on the test, Robbin does give the resulting formula as:

$$U = \frac{1}{1 - t_u}\left[(\text{Target ROE})\left(\frac{\text{E:S Ratio}}{\text{P:S Ratio}}\right) - i_{AT}\left(PHSF + \frac{1}{\text{P:S Ratio}}\right)\right]$$

Advantages and Disadvantages

The CY ROE method is appealing because most of the inputs are readily obtained from reported financial statements and exhibits and it produces a metric that is similar to the GAAP ROE commonly used to measure profitability in other industries.

However, like all calendar year methods it is subject to distortion during periods of growth and it requires a target rate of return be established and defended. It also requires a premium to surplus ratio, which may have to vary by line and which may not necessarily balance to actual surplus. Also, it uses GAAP equity in the denominator, which may be debatable, but this is done to ensure consistency with the GAAP ROE calculation typically performed using actual data (and the likely source for helping to establish an appropriate target ROE).

IRR on Equity Flows

A complete discussion of this method was shown in the previous Section of these notes, along with the Feldblum IRR paper. For continuity, I'll simply summarize those notes here.

The IRR method tracks the equity flows that would occur had we set up a company to write a *single* policy. Setting the value of the initial investment equal to the present value of the future equity flows to the hypothetical shareholders allows us to calculate an internal rate of return (IRR). Setting that IRR equal to a target IRR then allows us (with a fair amount of work if done by hand) to solve for the underwriting profit margin needed.

The Equity Flows

At inception there is some amount of "required surplus", which induces a negative flow for the shareholders (to the company). At the end of each period, we assume that the company pays out all of its income to the shareholders, a positive equity flow, but also has to maintain an adequate amount of surplus to meet statutory and/or GAAP requirements, which might result in additional equity flows, positive or negative, in that period. That is, any required increases in surplus require more flows from the shareholders to the company and any decreases result in equity flows to the shareholders.

Key Formula

$$0 = \sum_{j=0}^{n} \frac{\left[INC_j - SCHNG_j\right]}{(1 + IRR)^j}$$

Notice that the summation index runs from $j = 0$ because all of the flows are assumed to occur at the beginning of each period. That means that INC_0 will reflect any underwriting income or loss at inception (limited usually to the acquisition expenses incurred since no premium is earned yet). Similarly, $SCHNG_0$ will reflect the required surplus at inception.

Advantages and Disadvantages

An advantage of this approach is that the IRR is relatively simple to interpret — it reflects the expected "interest" earned on the shareholders' "loan" of surplus to the company. It also nicely captures the impact that accounting rules and regulatory capital restrictions directly impact shareholders.

However, to use this we still need to assume a particular surplus requirement, at inception and over time, which may not make sense on a line of business or policy basis. We also need to identify a target IRR. And then that forces us to confront the overall challenge that riskier lines may need more surplus, may require a higher IRR or maybe both. Having to set both is therefore tricky.

PV of Income to PV of Equity Ratio

This method calculates a ratio of the present value of income to the present value of average equity balances. It doesn't really produce a rate of return in the usual fashion, but the ratio is usually interpreted as a rate of return and used to compare against a target rate. Setting the ratio equal to the target return, we can then solve for the value of U.

The key point to note about this method, in addition to the specific way in which the inputs are determined, is the specific method used to calculate the present values.

Key Formula

The key formula is:

$$\text{Target Return} = \frac{\text{Present Value of Income Stream}}{\text{Present Value of Avg Equity Balance}}$$

$$r = \frac{PV(INC)}{PV(EQB)}$$

Calculation Details

The calculation of the income stream, INC, is identical to what we used in the IRR method. To reiterate, we follow Statutory and GAAP accounting rules to develop income statements and balance sheets for a hypothetical company writing the policy or line of business under review. We follow all of the relevant rules, including calculating taxes in accordance with IRS guidelines (e.g. reserve discounting).

Usually when we calculate a rate of return we have an *end of period* value in the numerator and a *beginning of period* value in the denominator. Robbin suggests rectifying this by discounting the income flows to the end of the first year rather than to the beginning of the first period.

For the equity balances, notice that unlike in the IRR calculation where we were concerned with the flow of equity funds to the shareholder, here we want to capture the stream of average equity balances. For this purpose, we set the equity balance each period equal to the average balance during the year:

$$EQB_i = \frac{EQ_{i-1} + EQ_i}{2}$$

When calculating the present value of this series of equity balances, it would seem natural to treat these balances as being committed at the *beginning* of the period so that the first value (EQB_1) is discounted for zero periods, the second balance is discounted for one period, etc. But in Robbin's first formula for this calculation, he uses an extra period of discounting, calculating the present value as follows:

$$PV(EQB) = \sum_{j=1} (EQB_j)^j$$

Whether he really intended it to be this way is unclear. In other papers and presentations he has given on this method this was not the case and instead he used:

$$PV(EQB) = \sum_{j=1} (EQB_j)^{j-1}$$

However, as will be discussed below, he also applies an adjustment to account for the effect of using quarterly reserve balances in this calculation. Because he also uses an extra period of discounting in this adjustment, the result is the same as if the discounting were done without the extra period. So in the numerical problems I will follow Robbin's numerical example and include the extra period of discounting as well.

Effective Annualized PVI/PVE

When using quarterly equity balance figures, an additional adjustment is needed. Notice that the denominator of the PVI/PVE ratio reflects the discounted sum of all the balances. That means that if we used this in the case where the equity is held constant over a year, say at \$100, but calculated the ratio on a quarterly basis, with a 10% discount rate the denominator would be \$377. Here we invest \$100 for one year and we divide the discounted sum of the income stream by almost four times our investment? That's odd. It might be better to annualize the denominator by dividing by the sum of the quarterly present value factors.

The resulting annualized formula, including the effect of adjusting the numerator so that discounting is done to the end of the first year, is:

$$PVI/PVE_{Ann} = \frac{(1+i)\sum_{j=1}(INC_j)v^j}{\frac{\sum_{j=1}(EQB_j)v^j}{\sum_{j=1} v^j}}$$

Discount Rates

Robbin suggests using current pre-tax, risk-free rates to discount the income stream and including any tax payments in the income stream. He suggests that the same rate should be used for the equity balance discount rate too.

However, he also states that using the target rate of return can be justified as well. In other papers he has written that discuss this method he advocated using the target rate, assumed to be the company's cost of capital, for discounting both the income and the equity balances. His rationale is that this rate reflects that rate at which the company can borrow funds, giving the present value figures a meaningful interpretation.

Advantages and Disadvantages

This method produces a measure of return, albeit a strange one, that is not distorted by historical balance sheet growth rates and is somewhat comparable to a GAAP ROE in that it measures the ratio of income to equity. It can be thought of as a multi-year extension of the GAAP ROE measure that reflects the long-term nature of the equity commitment. However, it does require selection of discount rates, selection of a target ratio and an assumption regarding required equity capital in each period.

PV Cash Flow Return

To attract equity capital to an insurance company in support of writing a policy, the present value of the investors' equity flows must be equal to (or less than) the present value of the after-tax cash flows from underwriting and investment. Here, we include investment income on the invested surplus within our cash flows, but do not explicitly add in the investment income from policyholder supplied funds because these get reflected through the discounting mechanism.

In this approach, Robbin discounts the equity flows at the cost of capital, or the target rate of return, and the underwriting and investment income cash flows at the investment yield.

Key Formula

We use the following equation to solve for the value of U:

$$PV(\Delta\,\text{Equity; Target Return}) = PV(\text{Total Cash Flow; Investment Yield})$$

$$PV(\Delta EQ; r) = PV(TCF; i)$$

Key Inputs

The after-tax cash flows should, like the other methods discussed, take into account actual tax computations. However, Robbin suggests an approximation which applies the tax rate to the present value of cash flows and discounts the equity flows at an after-tax investment yield.

The investment income included here should reflect any adjustments to surplus that might be needed to reflect actual invested assets.

Advantages and Disadvantages

The focus on the present value of underwriting cash flows is appealing, but there is no easy way to reconcile this to a GAAP ROE measure because of differences in timing between cash flows and GAAP income.

Risk-Adjusted Discounted Cash Flow Method

This method simply calculates the fair premium as being equal to the present value of all loss and expense cash flows, including taxes paid on the present value underwriting income and taxes paid on the investment income on surplus.

The underwriting profit margin is then implied by the difference between this premium and the undiscounted expected losses and expenses.

Discounting Details

A key point to note about this method is that the discounting is done using *risk-adjusted* discount rates. In the paper, Robbin actually uses risk-free rates for all cash flows other than the loss cash flows, but this is just a simplification in the paper. It does not have to be done this way, assuming one could justify an appropriate risk-adjusted rate for those other cash flows.

In addition, in the numerical example in the paper the discounting is done *to the end of the first year*. This last point isn't obvious from the description and seems to be ignored on old exam questions. In addition, in other papers and presentations Robbin has given on this method, he ignored this last point as well. You should be aware of how it is done in the paper, but I think discounting to time zero is likely to be fine — and much clearer to everyone.

Federal Income Tax Term

A confusing aspect of the presentation of this method is the composition of the FIT term. It includes the taxes on underwriting income and the taxes on the investment income of the surplus. But what about the tax on the other investment income from policyholder supplied funds?

The investment income on policyholder supplied funds is actually included implicitly when we discount the losses and expenses. Therefore, when we apply the tax rate to this *discounted* underwriting income we capture the taxes on the investment income from policyholder supplied funds. We'll see this below.

Key Formula

Denoting the risk free rate of return as i_f and the risk-adjusted rate of return i_r we can write the key formula for this method as:

$$PV(\text{Premium}; i_f) = PV(\text{Loss}; i_r) + PV(\text{Expenses}; i_f) + PV(\text{Taxes}; i_f)$$

$$PV(P; i_f) = PV(L; i_r) + PV(FX; i_f) + PV(FIT; i_f)$$

Notice that this is a present value premium calculation, so it reflects the timing of the premium payments.

Alternative Formula

In the main presentation of this method the taxes (on both underwriting income and investment income on surplus) were treated as a combined item and labeled FIT. We can approximate this result using a slightly different form of the above equation where the tax on underwriting income is reflected by simply multiplying the present value underwriting income $(PV(P) - PV(L) - PV(FX))$ by the tax rate, t, and then separately including the present value tax on investment income on surplus.

Given this, we can use the following approximation:

$$PV(P;i_f) = PV(L;i_r) + PV(FX;i_f) + t[PV(P;i_f) - PV(L;i_r) - PV(FX;i_f)] + tPV(\text{Investment Income on Surplus})$$

Combining some of the terms we can rewrite this as:

$$PV(P;i_f)(1-t) = (1-t)PV(L;i_r) + (1-t)PV(FX;i_f) + tPV(\text{Inv. Income on Surplus})$$

Dividing through by $(1-t)$ we can write this as:

$$PV(P;i_f) = PV(L;i_r) + PV(FX;i_f) + \frac{tPV(\text{Inv. Income on Surplus})}{(1-t)}$$

This form of the equation is easier to apply sometimes depending on the information provided.

Risk-Adjusted Discount Rate

The critical aspect of this method is the risk-adjusted discount rate used to discount the losses. Robbin essentially uses the CAPM. Using his notation:

$$i_r = i_f + \beta[i_m - i_f]$$

However, he has to adapt this to be useful for reflecting the systematic risk in the *liability* cash flows. One way to do this is to estimate the equity betas for insurance stocks and recognize that the beta of the invested asset portfolio must equal the weighted average betas of the liabilities and the firm's equity (from the basic identity, assets equals liabilities plus equity).

For typical P&C company equity betas and invested asset betas, the above approach is likely to lead to a reasonable result in that the beta will be *negative*. This is precisely what we would expect because it will lead to positive risk margins for insurers assuming liability risk. That is, the discounted loss amount will be greater than the risk-free present value of the liability.

However, this is difficult to apply in practice because it is hard to infer line of business specific betas[1].

Be sure to understand the reason why we would expect the insurance premium to reflect discounted losses at a rate *below* the risk-free rate. Recall that for the policyholder, the claims that they will make against their insurance policy will be perfectly negatively correlated with the payments they will have to pay in loss scenarios. As a result, they rightly should be willing to pay a positive risk premium to hedge this risk. We want to be paid a risk premium for assuming risks that are positively correlated with our existing exposures, so it only makes sense

[1] Cummins and Phillips have used a technique known as the Full Information Beta approach. They apply it to derive both the CAPM beta and the Fama-French 3-Factor model betas for the industry that vary by line of business.

that we should expect to pay risk premiums for transactions that are negatively correlated with our existing exposures.

Advantages and Disadvantages

This method has a somewhat solid theoretical foundation (though this could be debated), which perhaps could be enough of a positive to live with the uncertainty in estimating liability betas, especially by line of business. Other advantages are that it doesn't require a target return or an allocation of surplus to implement.

Practice Questions

Question 1. Three methods presented in the Robbin reading (CY Offset, PV Offset and CY ROE) require the company's after-tax investment yield. Describe the process that Robbin recommends to calculate this using reported financial data for the most recent calendar year.

Solution. In these calculations we always want the prospective after-tax yield, which can be estimated using actual dividend and interest yields and realized capital gains for the most recent period. New money rates, or current yields on bond investments, are more theoretically sound, but determining expected returns including capital gains on bonds or expected returns on equities, real estate, etc. can be very difficult to do and to justify.

There are two points with respect to this calculation using the company's actual returns by asset class.

First, he recommends using effective tax rates that vary by asset class to take into account factors such as the dividends received deduction, which reduces the tax rate on dividends from certain equity investments.

Second, he recommends including only realized capital gains and not unrealized capital gains.

Question 2. An insurer is planning to price their business using an underwriting profit margin equal to $U = 2\%$. Assume that on a calendar year basis the total policyholder supplied funds are 110% of premium, that the premium to surplus ratio is 2.0, the after-tax investment yield is 4.00% and the effective corporate tax rate on underwriting income (or loss) is 35%. Also assume that GAAP Equity and Statutory Surplus are equal to each other. Calculate the expected CY ROE and the underwriting profit margin that would be needed to achieve a 20% expected CY ROE.

Solution. With the information given and the CY ROE method, we can calculate the expected CY ROE by taking the sum of the after-tax investment yield on both policyholder supplied funds ($PHSF$) and surplus and the after-tax underwriting profit. For simplicity, we will assume $P = \$1$. And recall that we don't have to distinguish between GAAP Equity and Statutory Surplus here, so the formulas are somewhat simplified.

You can use the exact formulas shown in the earlier explanation, but it might be helpful to see the results using slightly different formulas because you need to be able to calculate the components using a wide range of inputs. Here, we want the total income relative to surplus, so we can actually start with just the sum of the after-tax underwriting profit margin and the after-tax investment income on the policyholder supplied funds ($PHSF$) as a percent of premium. That quantity isn't our final answer because it is relative to premium rather than to surplus and it doesn't include investment income on surplus. But to get there, we just multiply by the premium to surplus ratio (P:S Ratio) and then add in the after-tax investment yield:

$$\begin{aligned} ROE &= [(1 - t_u) \cdot U + i_{AT} \cdot (PHSF)](\text{P:S Ratio}) + i_{AT} \\ &= [(1 - .35)(2\%) + 4\%(110\%)](2.0) + 4\% \\ &= 15.40\% \end{aligned}$$

That is the ROE at the planned underwriting profit margin. But if we wanted to target $ROE = 20\%$, we could just rearrange that equation and solve for U:

$$\begin{aligned} U &= \frac{ROE_{\text{target}} - i_{AT} - i_{AT}(PHSF)(\text{P:S Ratio})}{(1 - t_u)(\text{P:S Ratio})} \\ &= \frac{20\% - 4\% - 4\%(110\%)(2)}{(1 - 35\%)(2)} \\ &= 5.54\% \end{aligned}$$

Question 3. Suppose you are using the PVI/PVE method and were given the following GAAP income figures and GAAP equity balances over the next four quarters. Using an 8% annual discount rate, determine the ratio PVI/PVE (include Robbin's quarterly adjustment).

Quarter End	UW Income	Investment Income	Pre-Tax Net Income	After-Tax Net Income	Ending Equity
0	-15.00	0.00	-15.00	-9.75	60.00
1	5.00	1.00	6.00	3.90	45.00
2	5.00	1.00	6.00	3.90	30.00
3	5.00	1.00	6.00	3.90	15.00
4	5.00	1.00	6.00	3.90	0.00

Solution. The first step is to calculate the PVI. You were given several streams of income, but the one you need to use is the after-tax net income which includes investment income.

You can use your calculators to do this by first entering all of the cash flows (see the IRR example in the notes for the instructions), then specifying the *quarterly* 2% discount rate and finally computing the present value to get $PVI = 5.10$ at time zero, as shown below:

Quarter End	UW Income	Investment Income	Pre-Tax Net Income	After-Tax Net Income	PV Factor	PVI
0	-15.00	0.00	-15.00	-9.75	1.000	-9.750
1	5.00	1.00	6.00	3.90	0.980	3.824
2	5.00	1.00	6.00	3.90	0.961	3.749
3	5.00	1.00	6.00	3.90	0.942	3.675
4	5.00	1.00	6.00	3.90	0.924	3.603
					Total PVI	5.100

Notice that this value for PVI reflects discounting to time zero. Below we will adjust this to reflect the present value at the end of the first year.

The next step is to calculate the PVE, but before doing that we need to calculate the average equity balance for each period. Then we simply calculate the present value using the same discount factors as before:

Quarter End	Ending Equity	Average Equity	PV Factor	PV(EQB)
0	60.00			
1	45.00	52.50	0.980	51.47
2	30.00	37.50	0.961	36.04
3	15.00	22.50	0.942	21.20
4	0.00	7.50	0.924	6.93
		Total	3.808	115.65

In this table, when the equity balances are discounted they sum to 115.65. Because of the use of quarterly data this figure is distorted, so we make an adjustment by dividing by the sum of the first four quarterly discount factors, $\sum_{j=1}^{4} 1.02^{-j} = 3.808$.

The resulting ratio of PVI to PVE, including adjusting the numerator to reflect discounting to the end of the first year, is:

$$PVI/PVE_{Ann} = \frac{(1.08)(5.100)}{115.65/3.808} = 18.1\%$$

Note: As mentioned in the notes, it is not technically correct to discount the equity balances as shown in the table above since it reflects an extra quarter of discounting for each value. It is being done here solely to be consistent with the numerical example in the reading. In this case because we used the same discount factors for the quarterly adjustment calculation the end result is the same. Differences would arise, however, if the data provided were annual rather than quarterly.

Question 4. Some discounted cash flow and IRR models used by actuaries focus solely on the underwriting cash flows of premium, loss and expense. Why does Robbin advocate focusing on cash flows to equity holders rather than just policyholder cash flows.

Solution. These cash flows don't take into account the impact of statutory or GAAP accounting conventions which impact shareholders, such as the need to carry reserves, to recognize deferred acquisition costs under GAAP, etc. They also don't take into account surplus requirements.

Question 5. You are trying to determine the underwriting profit margin using Robbin's Risk-Adjusted Cash Flow method and are given the following summary information, which is less detailed than the example in his paper:

Expected Loss	65.00	
Expenses	15.00	
Pre-Tax Investment Portfolio Yield	5.0%	
Risk-Free Rate	4.0%	
Risk-Adjusted Rate	1.5%	
Tax Rate	34.0%	
	Pre-Tax Risk-Free Rate	Pre-Tax Risk-Adjusted Rate
PV Loss	60.32	63.78
PV Expenses	13.92	14.72
PV Investment Income on Surplus	1.54	1.58

In this table, all present value amounts are shown using both risk-free discounting and risk-adjusted discounting.

Determine a reasonable approximation for the underwriting profit margin in this case, assuming that the full premium is paid up front.

Solution. Working through this method, and some of the others, on the exam can be time consuming because of all of the calculations you typically have to do. This question tried to make it easier for you, but I gave you extraneous information to ensure that you knew which values to use in the calculations. More importantly, by simplifying the inputs I didn't give you any way to directly calculate the stream of income tax payments.

Recall that the basic equation for this method presented in the paper was:

$$PV(P; i_f) = PV(L; i_r) + PV(FX; i_f) + PV(FIT; i_f)$$

However, I also showed a different form of the equation as:

$$P = PV(L; i_r) + PV(FX; i_f) + \frac{tPV(\text{Investment Income on Surplus})}{1 - t}$$

Given the information provided in this question, this form is easier to use.

$$
\begin{aligned}
P &= PV(L; i_r) + PV(FX; i_f) + \frac{tPV(\text{Investment Income on Surplus})}{1-t} \\
&= 63.78 + 13.92 + (.34)(1.54)/(1 - .34) \\
&= 78.49
\end{aligned}
$$

The question asked for the underwriting profit margin, which we get from:

$$
\begin{aligned}
U &= \frac{P - L - FX}{P} \\
&= \frac{78.49 - 65 - 15}{78.49} \\
&= -2\%
\end{aligned}
$$

Question 6. You are told that a line of business was priced using Robbin's Present Value Offset method and the following facts are given. Assume that for both the reference line and the reviewed line loss payments are all made on a single day *n* years from the inception date.

Undiscounted Expected Loss	100,000.00
Fixed Expense	3,000.00
Variable Expense (% of Premium)	15%
Investment Yield	4.00%
Payment Date in Years - Reviewed Line	3.00
UPP for Reviewed Line	1.09%

What is the premium that should be charged for this line?

Solution. This question is intended to highlight a key aspect of this method with regard to the underwriting profit margin.

Recall that I had shown a formula for the premium under this method in terms of the underwriting profit provision for the reference line:

$$P = L[PV(X)] + FX + VX(P) + U^0(P) + L[1 - PV(X^0)]$$

But I didn't give you any information related to the reference line, so this cannot be used. Instead though, I actually told you the value of U for the line being priced, $U = 1.09\%$. This means we can calculate the premium quite easily:

$$
\begin{aligned}
P &= \frac{L + FX}{1 - VX - U} \\
&= \frac{103,000}{1 - .15 - .0109} \\
&= 122,751
\end{aligned}
$$

Notice the point I was trying to highlight. The underwriting profit provision *includes* the impact of discounting and as a result when you use it in the premium formula you use it with undiscounted losses.

Background: Risk Loads for Insurers and Reinsurers

Introduction

The Mango reading makes extensive use of formulas that were originally derived in a paper by Kreps (not the one on the syllabus) and a paper by Meyers. The key elements of those two papers are presented here to simplify the presentation of the Mango paper.

Kreps: Reinsurer Risk Loads from Marginal Surplus Requirements

Introduction

Kreps uses ruin theory to establish rather simple risk load formulas that are easy to implement and appear to be consistent with reinsurance market pricing in many instances.

He makes the simple assumption that when adding a single policy to an existing book of policies, which presumably is already supported by some amount of surplus, what matters is the amount of *additional* surplus required to maintain the same ruin probability. Once this marginal surplus is measured, the Kreps method boils down to simply assuming that the risk load needs to be sufficient so that, after investment income is taken into consideration, the insurer earns an adequate rate of return on the marginal surplus requirement.

Derivation

Regarding the formulas that Kreps presents, the logic and derivation is rather simple. The only messy part comes into play when he develops a very simple formula to apply his approach.

He assumes that the capital (surplus) contributed, V, is equal to some multiple of the standard deviation less the expected profit (in dollars), or

$$V = zS - R$$

where z is a multiple of the standard deviation which results in some target (small) probability of ruin, S is the standard deviation of the existing book of business and R is the expected profit of the existing book of business. Writing an additional policy, with expected profit (i.e. risk load) r, standard deviation σ and correlation with the existing book C, we can write the *new* surplus, overall standard deviation and expected profit as V', S' and $R' = R + r$. This gives the following for the marginal surplus required:

$$V' - V = z(S' - S) - r$$

Then, the risk load is simply a target return (over the risk free rate) on this surplus so that investing surplus to write this policy is at least as attractive as investing the surplus in other marketable securities with similar risk.

Denoting this target return as y we have:

$$\begin{aligned} r &= y(V' - V) \\ &= y[z(S' - S) - r] \\ &= \frac{yz}{1+y}(S' - S) \end{aligned}$$

With that, the rest of the Kreps formulation is really just simplification of the notation. The messiest step is to notice that $(S' - S)$ can be written in terms of the stand-alone standard deviations and correlation as follows:

$$\begin{aligned} S'^2 &= S^2 + \sigma^2 + 2S\sigma C \\ S'^2 - S^2 &= \sigma(2SC + \sigma) \\ (S' - S)(S' + S) &= \sigma(2SC + \sigma) \\ (S' - S) &= \frac{\sigma(2SC + \sigma)}{S' + S} \end{aligned}$$

Later we will show an approximation for this quantity when the new risk is very small compared to the existing book. But first, let's see how Kreps simplifies his end result. I will begin with the result shown earlier and then plug in the value just derived for $(S' - S)$.

$$\begin{aligned} r &= \frac{yz}{1+y}(S' - S) \\ &= \frac{yz}{1+y}\frac{\sigma(2SC + \sigma)}{S' + S} \\ &= \left[\frac{yz}{1+y}\frac{(2SC + \sigma)}{(S' + S)}\right]\sigma \\ &= \Re\sigma \end{aligned}$$

where $\Re$ is the symbol used to reflect the quantity within the brackets. It is what Kreps calls his *reluctance* factor. It simplifies Kreps' approach down to a simple standard-deviation based risk load.

Well, sort of. There is still a σ term inside this reluctance term and so it may not be entirely fair to present Kreps' risk load as a simple multiple of the stand-alone standard deviation. In any event, this reluctance factor is appealing because it embodies all of the characteristics we would want it to reflect:

- it is an increasing function of the target return, y
- it is a decreasing function of the probability of ruin, since as the target probability of ruin increases, z falls and $\Re$ rises
- it is an increasing function of the correlation of the new policy to the existing book

Special Case

Consider the special case where the new policy is very small compared to the existing book. In this case $S' \approx S$ and therefore,

$$\Re \approx \frac{yz}{1+y}\left[C + \frac{\sigma}{2S}\right]$$

Reluctance Factors in Practice

Using the Kreps formulation for for the reluctance factor, at the time of the paper Kreps noted that this tends to range from .30 to .70 in practice.

Meyers: Marginal Variance Approach

You do not need to know much about the Meyers approach, other than to be familiar with the formulas used by Mango in his paper.

First, Meyers developed a risk load formula for a new risk, n, being added to an existing portfolio, L, of the form:

$$\begin{aligned} r &= \lambda[\text{Marginal Variance of Adding } n \text{ to } L] \\ &= \lambda[\text{Var}(n) + 2\text{Cov}(n, L)] \end{aligned}$$

Like the Kreps result, the risk load depends on both the variance of the stand-alone risk as well as its covariance with the existing portfolio. Unlike the Kreps result though, which had an intuitive and simple formula for the multiplier on the marginal surplus, Meyers' λ multiplier is less intuitive and harder to quantify.

You don't need to know anything about how Meyers advocated developing his lambda. However, you should realize that Mango attempts to show a consistent result for the Kreps and Mango methods and so he converts from the Kreps multiplier $yz/(1+y)$ to the Meyers λ marginal variance multiplier by dividing it by the standard deviation of the combined portfolio containing both L and n.

$$\lambda = \frac{yz/(1+y)}{\text{Std Dev}(L+n)}$$

Mango: Property Catastrophe Risk Load

Introduction

In this paper Mango shows that two popular risk load methods, Kreps' Marginal Surplus (MS) method and Meyers' Marginal Variance (MV) method suffer from the problem that a given risk's risk load depends upon the order in which it is added into an overall portfolio. That is, if we view the risk as the *first* one written it will command a different risk load than if it was considered the *last* one written.

Mango points out that this is particularly troublesome for setting renewal pricing because if we take a given portfolio and then sequentially calculate the renewal risk loads for each of the policies as if all of the other policies were written (that the policy in question is always the *last* one written, then our combined risk loads will either be less than risk load required for the whole portfolio (in the case of the MS method) or greater than the risk load required for the whole portfolio (in the case of the MV method).

Mango shows that this boils down to how the different methods allocate the covariance between a given policy and all of the other policies in the portfolio and proposes using a more general framework that explicitly allocated the covariance in a "fairer" manner. He uses property-catastrophe reinsurance for his examples in part because it makes it easier to calculate the covariance if all of the policy outcomes are specified in a property catastrophe event loss table applicable to all policies.

Review of Kreps and Meyers

I provided a more detailed discussion of these two methods in an earlier reading, so I will be very brief here. Mango uses two risk load methods:

- Marginal Surplus — Using the Kreps paper, and denoting the target return on allocated capital as y, the percentile at which the required capital is set as z, S as the standard deviation of the existing portfolio and S' as the standard deviation of the portfolio including the new policy, the marginal surplus risk load formula for a new risk being added to an existing portfolio is:

  $$\text{MS Risk Load} = \frac{yz}{1+y}(S' - S)$$

  The quantity in the brackets is the marginal standard deviation of the new risk and the term $yz/(1+y)$ is a multiplier.

- Marginal Variance — Using Meyers' paper, the risk load for a new risk n being added to an existing portfolio L is some multiple λ multiplied by the marginal variance:

  $$\text{MV Risk Load} = \lambda[\text{Var}(n) + 2\text{Cov}(L, n)]$$

In this case the parameter λ is not as clearly defined, but Mango uses the MS multiple, $yz/(1+y)$, scaled by the standard deviation of the combined portfolio:

$$\lambda = \frac{yz/(1+y)}{\text{Standard Deviation of } L+n}$$

Numerical Example

Let's be sure we know how to apply the above two formulas.

For this example we will assume that we have two risks, X and Y, and that we want to sequentially write them, starting with X and then writing Y. The stand-alone and combined standard deviations and variances, as well as their covariance, are given in Table 1:

TABLE 1. Summary of Variance Assumptions for X and Y

	Stand-Alone X	Stand-Alone Y	Combined
Standard Deviation	4,429	615	4,785
Variance	19,619,900	377,959	22,898,959
Covariance			1,450,550

We will further assume that under the Kreps method the target return on marginal surplus is 20% and the required capital is calculated at the 97.725th percentile, which corresponds to a *z-value* of $z = 2.0$.

Marginal Surplus Method

Under this method, we need to know the marginal standard deviations and the risk load multiplier. The latter figure is given as:

$$\text{Risk Load Multiplier} = \frac{yz}{(1+y)} = \frac{.2(2)}{1.2} = .33$$

For the marginal standard deviations, we assume we start by writing X, so it's marginal standard deviation is just its stand-alone standard deviation, or $4,429$ and therefore its risk load is $1,461.57$.

If we now add Y, we have the marginal standard deviation equal to:

$$S' - S = 4,785 - 4,429 = 356$$

Hence, the risk load is simply $356 * .33 = 117.48$.

Notice that if we applied the same risk load formula to the complete portfolio, its risk load would be $4,785 * .33 = 1,579$ which is the sum of the risk loads for X and Y.

However, consider what would have happened if we assumed we would start with Y and then add X? Here, we would still get the correct total risk load of $1,579$, but we would have had very different values for the respective risk load:

$$\text{Risk Load } Y = .33(615) = 202.95$$

$$\text{Risk Load } X = .33(4,785 - 615) = 1,376.10$$

This shows that the risk loads using this method depend on the order in which the policies are added to the portfolio.

Marginal Variance Method

The steps are essentially the same, but first we need an appropriate multiplier, which we will get using Mango's adjustment to the MS multiplier:

$$\lambda = \frac{yz/(1+y)}{\text{Standard Deviation of } L+n} = \frac{.33}{4,785} = .000069$$

Now, the marginal variance for X is again just the stand-alone variance since we are assuming it is added first. And the marginal variance for Y is the difference between the total variance and the variance for X, the existing portfolio. The resulting risk loads are:

$$\text{MV Risk Load } X = (.000069) * 19,619,900 = 1,353.77$$

$$\text{MV Risk Load } Y = (.000069) * (22,898,959 - 19,619,900) = 226.26$$

Notice that in this case, we could have also calculated the marginal variance for Y as:

$$\text{Marginal Variance } Y = \text{Var}(Y) + 2\text{Cov}(X,Y)$$

And the risk load as:

$$\begin{aligned}\text{Risk Load } Y &= .000069[\text{Var}(Y) + 2\text{Cov}(X,Y)]\\ &= .000069[377,959 + 2(1,450,550)]\\ &= 226.25\end{aligned}$$

And again, notice that had we started with Y and added X, we would have produced the same total risk load, 1,580.03 but the amounts for each policy would have been different. In this case, the risk loads would be:

$$\text{MV Risk Load } Y = (.000069) * 377,959 = 26.08$$

$$\text{MV Risk Load } X = (.000069) * (22,898,959 - 377,959) = 1,553.95$$

Renewing Policies

I noted in the numerical examples that the risk loads for each transaction depend on the order in which they occur. This poses a problem for a steady-state portfolio that renews each of its policies, for instance, simultaneously. It would clearly be wrong to treat each policy as if it was the *first* policy written. But it would also be a problem to treat each policy as if all the others were written and it was the *last* policy written. Let's see why using the results obtained above.

Example Using MS Method

Suppose that, upon renewal, we treated both X and Y as if each was the last policy written so that each can get the benefit of the diversification with the existing portfolio. This means that the risk load for X would be calculated using the MS method as if Y were already written and X was being added to it. In the case of the MS method, we saw this amount as $1,376.10$. And for Y, we would calculate the MS risk load as if X was already written and Y was being added to it, which would be 117.48.

But notice that the sum of these risk loads is now $1,493.58$, which is lower than the risk load for the combined portfolio.

Example Using MV Method

We have the same issue, albeit slightly differently, with the MV method. Here, treating each policy as if it was the last so that the pricing could reflect diversification would produce risk loads for X and Y of $1,553.95$ and 226.26. This adds up to $1,780.21$, which is now larger than the risk load under the MV method for the complete portfolio.

Renewal Additivity

What the previous examples show is that both the MS and MV methods suffer from the problem that the *renewal risk load*, calculated for a given risk as if it is the last risk in the portfolio and all others have been written, is not additive across all policies. In the case of the MS method the renewal risk load understates the total risk load needed. In the case of the MV method it overstates the total risk load needed.

To see why, consider only the marginal variance calculations that we did before, summarized in algebraic notations as:

$$\text{Total Variance} = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

Now if we look at the marginal variance calculated as in the renewal risk load calculation, whereby we treat each risk as if it is the last risk, then:

$$\text{Marginal Variance for} X = \text{Var}(X) + 2\text{Cov}(X, Y)$$

$$\text{Marginal Variance for} Y = \text{Var}(Y) + 2\text{Cov}(X, Y)$$

Since both of these marginal calculations include two times the covariance, the sum of the marginals will clearly add up to more than the total variance of the portfolio because the covariance terms will be double-counted.

The Shapley Value

The fact that the source of the problem, at least in the MV case, was the double-counting of the covariance term hints at a solution, which is to measure the renewal risk loads using an allocation of the covariance term in a way that doesn't double-count it. The so-called Shapley Value accomplishes this because it would treat the marginal risks as:

$$\text{Shapley Value for Marginal Risk from Renewing } X = \text{Var}(X) + \text{Cov}(X, Y)$$

$$\text{Shapley Value for Marginal Risk from Renewing } Y = \text{Var}(Y) + \text{Cov}(X, Y)$$

Notice that now the covariance terms are not double counted because each of the policies X and Y share their covariance terms equally.

You aren't responsible for the background, but it is good to keep in mind that this Shapley Value is actually the elegant solution to a more brute-force calculation which is to calculate the marginal risk loads under the assumption that any given risk is the first, second, third, . . . , last risk in the portfolio and then averaging the results.

A More General Result — Generalized Covariance Sharing Method

The Shapley Value referenced above does indeed solve the renewal additivity problem, but it isn't necessarily an appropriate or practical method in the case of reinsurance due to the fact that it doesn't fully reflect the potential for the scale of the transactions to differ significantly.

The Shapley Value requires each transaction to share its covariance equally with the rest of the portfolio. But if one transaction is very large and the other is very small, it may not be fair to require them to share the covariance equally. Instead, we might want to apply weights, say W_X and $W_Y = 1 - W_X$, to the covariance term so that the marginal risks are calculated as follows:

$$\text{Marginal Risk from Renewing } X = \text{Var}(X) + (W_X)2\text{Cov}(X, Y)$$

$$\text{Marginal Risk from Renewing } Y = \text{Var}(Y) + (1 - W_X)2\text{Cov}(X, Y)$$

Of course, if $W_X = W_Y = .50$ then this is identical to the Shapley value. But we could perhaps more fairly, in the reinsurance context, assign the weights in proportion to their expected values so that in the event that one policy is much larger than the other it will receive a larger proportion of the total covariance.

Mango calls this the *Covariance Share* method and applies it only in the case of the MV method.

Numerical Example of Shapley and Covariance Share Methods

Recall from the earlier numerical example that the covariance between X and Y was $1,450,550$. The Shapley method of determining the marginal variance would produce the following risk loads in the renewal case:

$$\begin{aligned}\text{Shapley Risk Load from Renewing } X &= \lambda[\mathrm{Var}(X) + \mathrm{Cov}(X,Y)]\\ &= (.000069)[19,619,900 + 1,450,550]\\ &= 1,453.86\\ \text{Shapley Risk Load from Renewing } Y &= \lambda[\mathrm{Var}(Y) + \mathrm{Cov}(X,Y)]\\ &= (.000069)[377,959 + 1,450,550]\\ &= 126.17\end{aligned}$$

Similarly, the Covariance Share method would assign weights in proportion to the expected losses. Technically Mango does this by simulated event rather than in the aggregate, but I will skip over that and address it only in the numerical problems below. For now, assume that the covariance share weights given are $W_X = 80.26\%$ and $W_Y = 19.74\%$. The risk loads from above would be adjusted accordingly:

$$\begin{aligned}\text{CS Risk Load from Renewing } X &= \lambda[\mathrm{Var}(X) + W_X(2)\mathrm{Cov}(X,Y)]\\ &= (.000069)[19,619,900 + 80.26\%(2)1,450,550]\\ &= 1,514.43\\ \text{CS Risk Load from Renewing } Y &= \lambda[\mathrm{Var}(Y) + W_Y(2)\mathrm{Cov}(X,Y)]\\ &= (.000069)[377,959 + 19.74\%(2)1,450,550]\\ &= 65.59\end{aligned}$$

Both the Shapley and the Covariance Share renewal risk loads sum to the original MV total risk load of 1,580.03. The difference between the two of them is simply a matter of fairness so that more of the allocated covariance is applied to the transaction with the larger losses (by event).

Practice Questions

Question 1. You are given the values below for the mean, standard deviation and variances for two reinsurance policies, A and B. If the covariance between the two policies is 914,255, the target return on marginal surplus is 15% and required capital is calculated at the 99th percentile, what would be the risk load for each policy using the Marginal Surplus method as discussed in Mango's paper? Use the *build-up* case where each risk is added sequentially to the portfolio. For convenience, note that $\Phi^{-1}(.99) = 2.326$.

TABLE 2. Assumptions

	A	B	Total
Mean	179	735	914
Std Dev	615	2,393	2,816
Variance	377,959	5,724,740	7,931,208

Solution. To apply the marginal surplus method, we can use the following formula where y is the required return on marginal surplus, z is the standard normal z-value at the target percentile, S is the standard deviation of the portfolio before adding the risk and S' is the standard deviation of the portfolio after adding the risk.

$$\text{MS Risk Load} = \frac{yz}{1+y}(S' - S)$$

Start with policy A. Since this is the first one in the portfolio, its marginal standard deviation is the same as its stand-along standard deviation. Plugging in the values:

$$\text{MS Risk Load for } A = \frac{.15(2.326)}{1.15}(615) = 186.6$$

For B, we do the same but now the marginal standard deviation is the difference between the complete portfolio standard deviation, $S' = 2,816$ and the original standard deviation before adding B, or $S = 615$. The risk load is then:

$$\text{MS Risk Load for } B = \frac{.15(2.326)}{1.15}(2,816 - 615) = 667.8$$

Question 2. How does the sum of the risk loads in the previous question compare to the risk load for the total portfolio? Assume that the risk load for the portfolio is calculated using the MS method and that both policies are written at the same time.

Solution. Here, we could calculate the MS risk load on the whole portfolio as:

$$\text{MS Risk Load for Portfolio} = \frac{.15(2.326)}{1.15}(2,816) = 854.4$$

Notice that this is identical to the sum of the two risk loads in the previous question, 186.6 + 667.8 = 854.4

Question 3. Suppose that after writing the two policies in the previous two questions, each comes up for renewal. The intent is to renew each policy, so it is no longer appropriate to treat A as being the only policy in the portfolio, as we did before in the build-up case. What would be the renewal risk loads in this case, again using the MS method?

Solution. The renewal risk load is simply the risk load that would be calculated if the policy in question were assumed to be the *last* policy added. In the previous question, we actually calculated the Risk Load for B under this same assumption, so the renewal risk load for B is the same as before, 668.

For A, before we calculated the risk load as if it were the first policy in the portfolio. To calculate the renewal risk load, we need to recalculate this as if it were the last policy added. The calculations are similar to what we did before for B.

$$\text{Renewal Risk Load for } A = \frac{.15(2.326)}{1.15}(2,816 - 2,393) = 129$$

Notice that if we charged the renewal risk loads for each policy we would charge only a total of $668 + 129 = 797$. This is *less than* the total risk load we know that we need for the whole portfolio, or 854.

Question 4. In the previous question we saw that the MS method results in renewal risk loads that are inadequate for the whole portfolio. Is this true of the MV (marginal variance) pricing method too? Show the results numerically to prove the point using the same approach to estimating the MV method's $lambda$ used by Mango and rounding to 7 decimal places.

Solution. In the case of the MV method, the sum of the renewal risk loads are greater than the risk load for the whole portfolio.

To see this numerically, we simply redo the calculations before but now use the MV method. In this case, the risk load multiplier, λ is calculated (in Mango's paper) as:

$$\begin{aligned}\lambda &= \frac{yz/(1+y)}{\text{Portfolio Standard Deviation}} \\ &= \frac{.3034}{2,816} \\ &= 0.0001077\end{aligned}$$

And now, the two renewal risk loads can be calculated as follows:

MV Renewal Risk Load for A		MV Renewal Risk Load for B	
Portfolio Variance	7,931,208	Portfolio Variance	7,931,208
Variance of B	5,724,740	Variance of A	377,959
Marginal Variance from A	2,206,468	Marginal Variance from B	7,553,249
λ	0.0001077	λ	0.0001077
MV Renewal Risk Load for A	**238**	MV Renewal Risk Load for B	**813**

Notice that in this case the sum of the renewal risk loads is $238 + 813 = 1,051$. Had the MV method been used on the entire portfolio the risk load would have been $(0.0001077)(7,931,208) = 854$, which is the same as we had before due to the way in which we calculated a consistent value for λ. So the MV method results in renewal risk loads that add up to more than the portfolio risk load. This is because of the double-counting of the contribution that the covariance makes to the total portfolio — each risk is charged that full amount upon renewal.

Question 5. The previous question showed that the MV method produces renewal risk loads that are too high, summing to more than the portfolio risk load needed. Show why this is the case using the formula, in symbols, for the portfolio variance and the two marginal variance calculations done on a renewal basis.

Solution. Recall that for a portfolio consisting of risks A and B, the total variance is:

$$\text{Total Portfolio Variance} = \text{Var}(A) + \text{Var}(B) + 2\ \text{Covariance}(A,B)$$

When calculating the renewal risk loads, the marginal variances are given as:

$$\begin{aligned}\text{Marginal Variance from A} &= \text{Total Portfolio Variance} - \text{Var}(B)\\ &= [\text{Var}(A) + \text{Var}(B) + 2\ \text{Covariance}(A,B)] - \text{Var}(B)\\ &= \text{Var}(A) + 2\ \text{Covariance}(A,B)\end{aligned}$$

And similarly,

$$\text{Marginal Variance from B} = \text{Var}(B) + 2\ \text{Covariance}(A,B)$$

So in this case, both of the marginal variances include the term 2 Covariance(A,B), which causes the covariances to be counted multiple times.

Question 6. How does the Shapley value for the marginal variance avoid the problem referenced in the previous question with respect to counting the covariance multiple times?

Solution. The Shapley value for the marginal variances are:

$$\text{Marginal Variance from A} = \text{Var}(A) + .5(2)\ \text{Covariance}(A,B)$$

$$\text{Marginal Variance from B} = \text{Var}(B) + .5(2)\ \text{Covariance}(A,B)$$

In this case, the sum of the marginal variances is equal to the total portfolio variance and so the renewal risk loads are indeed additive. In the words of Mango, it has the property of being *renewal additive.*

Question 7. The previous question suggests that *any* set of factors that add to 1.0 could be used in place of the 0.5 factors used above in order to maintain renewal additivity. Is this true? If so, what should be the basis of selecting these weights?

Solution. It is true that any weights could be used. Using an arbitrary set of weights, assuming they add to 1.0, turns the Shapley value into what Mango calls the generalized covariance share value. All such weights will be renewal additive, but since the scale of any two risks could vary substantially, it would be unfair to require a very small risk to share its covariance with a very large risk equally. Instead, more weight should be shifted to the larger risk.

Question 8. What specific method does Mango suggest to use to establish the weights?

Solution. He recommended using the relative loss amounts by event.

Notice that because he recommends applying the relative weighting by event and not to the overall distribution, it is important to be able to do the calculations using the event loss tables as the data source. The next questions will address this.

Question 9. In Mango's numerical examples he uses an event loss table with six events, their associated probabilities and their losses by event for two risks, X and Y, as shown below. Using this occurrence size of loss distribution, calculate the expected loss and variance for each risk as well as the covariance between the two risks.

TABLE 3. Occurrence Size of Loss Distribution

Event	$p(i)$	$1 - p(i)$	Loss for X	Loss for Y	Total Loss
1	2.0%	98.0%	25,000	200	25,200
2	1.0%	99.0%	15,000	500	15,500
3	3.0%	97.0%	10,000	3,000	13,000
4	3.0%	97.0%	8,000	1,000	9,000
5	1.0%	99.0%	5,000	2,000	7,000
6	2.0%	98.0%	2,500	1,500	4,000

Solution. The three key formulas you need are:

$$\text{Mean} = \sum p(i)L_i$$

$$\text{Variance} = \sum p(i)[1 - p(i)]L_i^2$$

$$\text{Covariance}(L, n) = \sum p(i)[1 - p(i)]L_i n_i$$

These calculations are shown in the following table:

TABLE 4. Calculation of Mean and Variance

	Mean		Variance	
Event	$p(i)X_i$	$p(i)Y_i$	$p(i)[1 - p(i)]X_i^2$	$p(i)[1 - p(i)]Y_i^2$
1	500	4	12,250,000	784
2	150	5	2,227,500	2,475
3	300	90	2,910,000	261,900
4	240	30	1,862,400	29,100
5	50	20	247,500	39,600
6	50	30	122,500	44,100
	1,290	**179**	**19,619,900**	**377,959**

For the covariance calculations, we have:

TABLE 5. Calculation of Covariance

Event	$p(i)[1-p(i)]X_iY_i$
1	98,000
2	74,250
3	873,000
4	232,800
5	99,000
6	73,500
	1,450,550

Question 10. Use Mango's covariance share method and his proposed loss weighting to develop the Covariance Share Risk Load for the two risks for which we just calculated the mean, variance and covariance, assuming those are the only risks in the portfolio. For the risk load, assume $\lambda = .000069$

Solution. The point of this question is just to make sure you realize that the loss weighting is done *by event* and so calculating the covariance share follows the calculation of covariance but with an added term reflecting the weights given to each event.

The table below shows this calculation:

TABLE 6. Calculation of Covariance Share

Event	Loss for X	Loss for Y	Weight X	Weight Y	Cov Share X	Cov Share Y
1	25,000	200	99.2%	0.8%	194,444	1,556
2	15,000	500	96.8%	3.2%	143,710	4,790
3	10,000	3,000	76.9%	23.1%	1,343,077	402,923
4	8,000	1,000	88.9%	11.1%	413,867	51,733
5	5,000	2,000	71.4%	28.6%	141,429	56,571
6	2,500	1,500	62.5%	37.5%	91,875	55,125
					2,328,401	**572,699**

Notice that to be consistent with the exhibits in the table, the values in the table above reflect ***two times the event covariance*** *multiplied by the weights. For instance, in the previous question we found that the covariance for Event 1 was* $98,000$. *In the covariance share table, the covariance shares for Event 1 sum to* $2 * 98,000 = 196,000$, *with Cov Share X* $= 99.2\% * 2 * 98,000$. *This is an unfortunate inconsistency in the terminology, since it makes more sense to include the multiple of 2 in the risk load calculation rather than in the partitioning of the covariance.*

Then, to calculate the risk load, we use the following:

$$\text{Risk Load}(X) = \lambda[\text{Var}(X) + \text{CovShare}(X)]$$

These calculations for X and Y are shown below:

TABLE 7. Risk Load using Covariance Share Method

	X	Y
Variance	19,619,900	377,959
Cov Share	2,328,401	572,699
Total	21,948,301	950,658
Lambda	0.000069	0.000069
Risk Load	**1,514**	**66**

Question 11. In the previous question we had only two risks. Describe what you would do differently if there had been multiple risks in the portfolio.

Solution. When there are more than two risks in the portfolio you have to perform the above calculations for each *pair* of risks and then the shared covariance term (*CovShare*) for any one risk is the sum of the calculations across all risks in the portfolio.

Question 12. A reinsurer has three policies that it has written, X, Y and Z and it wants to use Mango's Covariance Share Method to determine the renewal risk loads for each. The following tables contain pertinent information needed. Use the data to develop the risk load for each risk using Mango's Covariance Share Method with $\lambda = .000069$.

TABLE 8. Portfolio Means and Variances

	X	Y	Z	Total
Mean	1,290	179	735	2,204
Variance	19,619,900	377,959	5,724,740	51,522,658

TABLE 9. Pairwise Covariance Share Calculations - Weighted by Event Loss

Covariance Share Calculations - Pairwise	
$\text{CovShare}^X(X,Y)$	2,328,401
$\text{CovShare}^X(X,Z)$	13,674,111
$\text{CovShare}^Y(Y,X)$	572,699
$\text{CovShare}^Y(Y,Z)$	513,001
$\text{CovShare}^Z(Z,X)$	7,396,339
$\text{CovShare}^Z(Z,Y)$	1,315,508
Total	25,800,059

Solution. The first step of the calculation is to get the Covariance Share in total for each risk. The table in the question gave the pairwise values, so all we need to do is calculate the sums across each pair. For instance,

$$\text{CovShare}^X = \text{CovShare}^X(X,Y) + \text{CovShare}^X(X,Z)$$

Then, using those, we add in their respective variances and multiply by λ. These calculations are shown below:

TABLE 10. Risk Load Calculations — X, Y, Z

	X	Y	Z
Variance	19,619,900	377,959	5,724,740
Cov Share	16,002,513	1,085,700	8,711,846
Total	35,622,413	1,463,659	14,436,586
Lambda	0.000069	0.000069	0.000069
Risk Load	**2,458**	**101**	**996**

Question 13. What calculations could you do to quickly verify that the covariance share risk loads in the previous question are indeed renewal additive?

Solution. The easiest way to do this is to calculate the risk load that would be appropriate for the whole portfolio and then confirm that the sum of the risk loads in the previous question is the same.

In the previous question you were told that the total portfolio variance was $51,522,658$. Applying the MV risk load method to the whole portfolio, the marginal variance for the whole portfolio is just the portfolio variance, so the total risk load would be:

$$\text{Portfolio Risk Load} = \lambda[\text{Portfolio Variance}] = 3,555$$

Quick summation of the risk loads found in the previous question confirms that these add up to $3,555$.

Question 14. Why does Mango say that the pricing of risk loads to maintain renewal additivity is really just an allocation problem?

Solution. Since we know what the total risk load is that we want to achieve, which is the MS, MV or some other pricing methods applied to the entire portfolio of risks, pricing an individual risk is really all about an allocation of this total. This can be achieved using a risk measure that incorporates a weighting of the covariance terms so that they add up to the total. Thus, it really comes down to a proper allocation of the quantity $2\text{Cov}(n,L)$ to each of the policies.

Kreps: Investment-Equivalent Reinsurance Pricing

Introduction

In this reading Kreps develops an algorithm and formulas for establishing a risk load for a stand-alone reinsurance transaction. He is careful to note that the risk loads are higher than they need to be because they do not in any way reflect the reality that in most cases the transaction in question will be part of a broader portfolio of risks.

Overview of Kreps' Approach

We know that when an insurer writes a policy it has to raise capital from shareholders (investors). In order to attract this capital, the shareholders' expected return on this investment has to be at least as good as the expected return of investing the same amount of capital directly in a portfolio of risky assets. In addition, we need to ensure that having taken on the insurance risk we invest the premium and shareholder capital funds into "safe" investments. Two definitions of safe investments are used, resulting in two different risk load calculations.

In addition, we need to raise sufficient capital from shareholders to ensure that the combined insurance and safe investment strategy is viewed as not being too risky from the perspective of both the policyholder and the shareholder.

From the policyholder perspective, we need to ensure that there are sufficient assets at the end of the period so that there is a high probability of paying losses up to some target loss level (the safety constraint). From the shareholder perspective, the combined insurance and investment strategy must have less variability than a direct investment of the same capital directly in a risky investment. Since both constraints must be satisfied, the amount of capital raised is the larger of the assets needed for each constraint.

Two Investment Strategies

Kreps considers two alternative strategies for ensuring that the assets are invested safely:

1. Swap Risk-Free Investment for Risky Investment — Suppose that the risk measure used suggests that \$1 million of capital were needed to support a reinsurance transaction. One approach to ensuring that this amount of assets is available to pay claims is to invest them in risk-free investments rather than risky investments. Doing so will incur (in some sense) an opportunity cost from the lost investment income, but will serve the desired purpose.

2. Purchase Put Options — An alternative is to allow the capital to be invested in risky assets earning expected returns greater than the risk-free rate, but to also use some of the capital to purchase put options on those assets, with a maturity coinciding with the maturity date of the liabilities and with the strike price being set at the level that

the assets would have grown to if invested at the risk free rate (which is referred to as the forward value of the assets).

We will use each of these two different strategies and require, in each case, that the internal rate of return (IRR) for the combination of the reinsurance transaction and the investment strategy be the same as what would have been achieved had the reinsurance transaction not been done and the assets invested at the risky investment rate, y.

This will result in two different sets of risk loads and Kreps then uses the *lower* of the two, arguing that it would be the most competitive and practical.

The Constraints

In terms of the amount of assets to be raised from shareholders, we need to ensure that two constraints are satisfied:

a. Safety Constraint — The safety constraint simply requires that the funds available to pay claims at the end of the year are at least equal to a specified loss safety level. This can be viewed as a constraint imposed by the policyholder.

b. Investment Variance Constraint — The variance constraint simply requires that the IRR of the combined reinsurance and investment strategy be no more volatile than a direct investment in the risky assets. This can be viewed as a constraint imposed by the shareholder.

In the case of these constraints, both have to be satisfied and so we use the more conservative of the two. That is, for each of the two strategies (Swap or Put Option) we calculate the risk loads using both of the constraints and then use the larger of the two. We then have a risk load amount for each of the strategies and we take as our final risk load the smallest of the two.

Formulas — Swap Strategy

In this strategy we plan to collect premiums, P, and invest our own assets, A, such that at the end of the year our risk free investment after paying the claim amounts, L, will grow to $(1+r_f)(P+A)-L$. Since our initial investment is A, the IRR of this combined reinsurance and financial strategy is given as:

$$(1+\mathrm{IRR})A = (1+r_f)(P+A) - L$$

where the total amount invested in the risk free assets is denoted as $F = P + A$. Since the premium consists of the discounted expected loss amounts plus the risk load, R, and we want the expected IRR to be at least equal to the investment return we could achieve by investing in risky assets, y, we can write this equation as follows and then solve for R:

$$(1+y)A = (1+r_f)\left[\frac{\mu_L}{1+r_f} + R + A\right] - \mu_L$$
$$\Rightarrow R = \frac{y - r_f}{1+r_f}A$$

Safety Constraint

We are not quite there yet because we don't know the quantity for A that will satisfy each of the constraints. One constraint is that we want the funds at the end of the year to be worth at least the loss safety level s (again, taken here to be a percentile of the loss distribution):

$$(1+r_f)(P+A) \geq s$$

This gives us a relationship between A and s that we can obtain by plugging into the previous equation:

$$A \geq \frac{s - \mu_L}{1+y}$$

Taking the value of A that is as small as it can be and plugging into the risk load formula from above gives us:

$$R = \frac{y - r_f}{1+r_f}\frac{s - \mu_L}{1+y}$$

Numerical Example

Consider a reinsurance transaction that has expected claim costs that are lognormally distributed with a mean of \$1 million and a standard deviation of \$2 million.

You want to use the swap strategy with the safety constraint to determine a risk load and premium for this risk. If you write this policy, you would want to have enough capital to ensure that, along with the premium, the invested funds will grow to at least the 99.9th percentile of the claim cost distribution, which based on the inputs here is \$22.548 million. Assume that your alternative use of capital would be to invest in a risky asset portfolio with a yield of 5.3% and a standard deviation of 8.4%. The risk free rate is 3.6%. Determine the risk load, the pure premium and the total premium assuming an expense ratio of 10%.

Let's go through the steps in the most logical fashion. Along the way we will introduce the notation used by Kreps for all of the relevant variables.

First, the loss safety level is the 99.9th percentile of the loss distribution. We get this by finding the lognormal distribution parameters μ and σ from the mean and coefficient of variation (CV = \$2 million/\$1 million = 2.0) as:

$$\sigma = \sqrt{\ln(1+\mathrm{CV}^2)} = \sqrt{\ln(1+(2.0)^2)} = 1.2686$$

$$\mu = \ln(\text{Mean}) - \sigma^2/2 = 13.0108$$

These parameters are then used to get the 99.9th percentile using the formula:

$$s = e^{\mu+N^{-1}(.999)\sigma} = e^{13.0108+3.09(1.2686)} = 22,548,347$$

Next, we determine the assets needed to satisfy the safety constraint as:

$$A = \frac{s-\mu_L}{1+y} = \frac{22,548,347-1,000,000}{1.053} = 20,463,767$$

With that, we can use the risk load formula:

$$\begin{aligned} R &= \frac{y-r_f}{1+r_f}A \\ &= \frac{.053-.036}{1.036}(20,463,767) \\ &= 335,795 \end{aligned}$$

This is the risk load only. To get the pure premium we add in the expected discounted claims, to get $P = 1,301,046$ and to get the total premium charged we gross that up for the 10% expense ratio to get $P' = 1,445,607$.

Variance Constraint

The previous result considered only the safety constraint. We also have another variance constraint which requires that the combined reinsurance and investment strategy have a dollar variance that is less than or equal to the dollar variance of the assets being invested directly in the target investment.

Using σ_L as the dollar loss standard deviation and σ_y as the standard deviation of the *rate of return*, the constraint is:

$$\begin{aligned} A\sigma_y &\geq \sigma_L \\ \Rightarrow A &\geq \frac{\sigma_L}{\sigma_y} \end{aligned}$$

At the equality, plugging in this for A gives a second risk load result of:

$$R = \frac{y-r_f}{1+r_f}\frac{\sigma_L}{\sigma_y}$$

The latter of these two is usually the higher, but in either case we use the one that is higher — for this financial strategy.

Numerical Example

Using the same facts as in the previous numerical example, calculate the risk load for the variance constraint and determine which risk load should be used for the swap strategy result.

Here, we know that the assets are equal to the ratio of the standard deviations of the claims, $\sigma_L = 2,000,000$, and the asset yield, $\sigma_y = 8.4\%$. This gives us:

$$\begin{aligned} A &= \frac{\sigma_L}{\sigma_y} \\ &= \frac{2,000,000}{.084} \\ &= 23,809,523 \end{aligned}$$

And then, as before, for this constraint:

$$\begin{aligned} R &= \frac{y - r_f}{1 + r_f} A \\ &= \frac{.053 - .036}{1.036}(23,809,523) \\ &= 390,697 \end{aligned}$$

Finally, we take the more conservative of the two risk load results, which in this case is the one based on the variance constraint and use that risk load as our selected risk load, or $R = 390,697$ under the swap strategy.

Subtle Point

One thing to be careful about is the assumption Kreps makes about the risky investment return. In his notation this has a mean value of y and a standard deviation of σ_y, but he also assumes that the return is lognormally distributed. Or to be more precise, that the quantity $(1 + y)$ is lognormally distributed so that it is a variable that can take on values from zero to infinity.

What this means is that when calculating the parameters of the distribution, μ and σ, we have to be very careful to convert the mean and standard deviation values in the same way that we did in the example above for the loss distribution so that we could calculate the 99.9th percentile.

The two key formulas you need to use for this are:

$$\sigma = \sqrt{\ln\left(1 + \left(\frac{\sigma_y}{1 + y}\right)^2\right)}$$

$$\mu = \ln(1 + y) - \frac{\sigma^2}{2}$$

We will use these formulas below when we have to calculate the value of a put option, but this point was made here because we will also need to use these formulas when we are dealing with a transaction whose cash flow occurs at some time, t, other than one year. In this case, we need to be sure to adjust our distribution parameters to μt and $\sigma\sqrt{t}$ and then recompute the compounded values of the mean yield, y and the yield standard deviation, σ_y.

Formulas — Put Option Strategy

The logic in this case is very similar, but the strategy is now a bit more complex. Now we are going to invest in the risky asset and purchase put protection so that we have no downside risk — the risky assets can always be sold for the amount needed to satisfy the loss safety level.

The put will require an actual cash outlay, impacting the safety constraint, and the investment in the risky assets will result in more volatility, impacting the variance constraint. But the real tricky part is that the downside risk is eliminated with the put, so the actual calculations of the safety constraint and the variance constraint are less straightforward than before.

As before, the funds available to be invested include the premiums charged (discounted expected claim payments plus risk load) and the assets supplied to satisfy the safety and variance constraints, but they are reduced by the cost of purchasing the put option. Denoting the price per dollar of notional value for the put as, r, we have the following for the total funds invested:

$$\begin{aligned} F &= P + A - rF \\ &= \frac{P + A}{1 + r} \end{aligned}$$

These funds will now grow at an uncertain rate with an expected value of i. This expected investment return differs from the yield on the risky assets, y, because it reflects the mean return on the hedged portfolio — that is, it reflects the elimination of the downside from the distribution.

As before, we can determine the IRR reflecting our investment of capital, A, the final value of the hedged investment portfolio and the paid claims. Setting this IRR equal to the investment rate on the risky assets so that the portfolio containing the reinsurance transaction and the hedged asset portfolio has the same mean return as if all of the capital were just invested directly in the risky assets, we have the following relationship between the total invested funds, F, and the amount of capital invested, A:

$$(1 + y)A = (1 + i)F - \mu_L$$

As a reminder that the risk load is buried within the value for F, note that we could also write this equation as a relationship between the risk load, R and the amount of capital invested, A:

$$\begin{aligned} (1 + y)A &= (1 + i)\frac{P + A}{1 + r} - \mu_L \\ &= (1 + i)\frac{R + \frac{\mu_L}{1 + r_f} + A}{1 + r} - \mu_L \end{aligned}$$

We can rearrange this in terms of R, which isn't critical but it is helpful to see it because this is the end result we are seeking. Simply multiply both sides by the ratio $(1+r)/(1+i)$ and simplify:

$$R = A\left[\frac{(1+r)(1+y)-(1+i)}{1+i}\right] + \mu_L\left[\frac{1+r}{1+i} - \frac{1}{1+r_f}\right]$$

Notice that we can't use this yet because we don't yet have a formula for A. To get that we have to apply each of the constraints regarding the safety level and the variance.

Safety Constraint

The safety constraint requires that whatever funds we invest, F, grow to be at least as big as the safety level, s. In the put option case, the funds will grow to *at least* $(1+r_f)F$ because we have a put that ensures this is the case.

So the safety constraint just says that this worst case must be greater than or equal to s, or:

$$(1+r_f)F \geq s$$

At the equality,

$$F = \frac{s}{1+r_f}$$

which can be plugged into the equation we had before for the relationship between A and F and then simplified:

$$(1+y)A = (1+i)F - \mu_L$$

$$(1+y)A = (1+i)\left[\frac{s}{1+r_f}\right] - \mu_L$$

$$A = \frac{1}{1+y}\left[\frac{1+i}{1+r_f}s - \mu_L\right]$$

Notice that this formula, unlike the similar formula in the swap case, now contains the variable i which reflects the mean return from the put-protected portfolio. This has to be calculated using very messy formulas, which I will not show here. But note that it reflects the expected return on the invested funds taking into account that they earn the risky asset return but are protected on the downside to grow at no less than the risk-free rate.

If we plug this formula for A back into the messy formula relating A and R we can solve for R as:

$$R = \frac{1}{(1+r_f)(1+y)}\left[s\left((1+y)(1+r)-(1+i)\right) - \mu_L(y-r_f)\right]$$

Numerical Example

Let's continue with our earlier numerical example and now apply the put option strategy along with the safety constraint. In this case, we again calculate the value for A at the equality, which is given as:

$$A = \frac{1}{1+y}\left[\frac{1+i}{1+r_f}s - \mu_L\right]$$

But now notice that we need the value for i. The formulas are in Kreps' appendix, which I won't repeat here, but when used the value for the expected hedged portfolio return is $i = 7.84\%$. Using this we get:

$$A = \frac{1}{1.053}\left[\frac{1.0784}{1.036}22,548,347 - 1,000,000\right] = 21,339,510$$

In a moment we will need the cost of the put option, as a proportion of the amount invested, which we get from the Black-Scholes formula as:

$$\begin{aligned} r &= \Phi\left(\frac{\sigma\sqrt{t}}{2}\right) - \Phi\left(\frac{-\sigma\sqrt{t}}{2}\right) \\ &= .0318 \approx .4\sigma\sqrt{t} \end{aligned}$$

Finally, plugging in to the messy formula for R as a function of A:

$$\begin{aligned} R &= A\left[\frac{(1+r)(1+y)-(1+i)}{1+i}\right] + \mu_L\left[\frac{1+r}{1+i} - \frac{1}{1+r_f}\right] \\ &= (21,339,510)\left[\frac{(1.0318)(1.053)-(1.0784)}{1.0784}\right] + 1,000,000\left[\frac{1.0318}{1.0784} - \frac{1}{1.036}\right] \\ &= 151,425 \end{aligned}$$

Notice that this would have been slightly less messy had we not first calculated A and instead just jumped to Kreps' final formula for R:

$$R = \frac{1}{(1+r_f)(1+y)}\left[s((1+y)(1+r)-(1+i)) - \mu_L(y - r_f)\right] = 151,425$$

Variance Constraint

Assuming that the investment return and the claims are uncorrelated, and denoting the standard deviations of the risky asset portfolio, the investment return (which is the net result of investing funds in the risky asset portfolio and buying put options) and the claims as σ_y, σ_i and σ_L, the variance constraint is:

$$(A\sigma_y)^2 \geq (F\sigma_i)^2 + (\sigma_L)^2$$

Again, we can substitute the formula for F in terms of A that we used earlier ($(1+y)A = (1+r_f)F - \mu_L$) to get a formula for A at the equality, but in this case it turns out that this formula is quite a mess. I will show it here for completeness though:

$$A = \frac{b + \sqrt{b^2 + ac}}{a}$$

where

$$a = \sigma_y^2(1+i)^2 - \sigma_i^2(1+y)^2$$
$$b = \mu_L(1+y)\sigma_i^2$$
$$c = \mu_L^2\sigma_i^2 + \sigma_L^2(1+i)^2$$

Once we know the value for A we can plug it into the formula we showed earlier for the risk load, R, in terms of A and μ_L. In this case, the formula for A is such a mess that I won't simplify the expressions any further. Just calculate A as above and plug it into:

$$R = A\left[\frac{(1+r)(1+y)-(1+i)}{1+i}\right] + \mu_L\left[\frac{1+r}{1+i} - \frac{1}{1+r_f}\right]$$

Numerical Example

Continuing with the same example, we will calculate the required capital, A, under the variance constraint in the put option strategy case. Here, we need to calculate one additional value, which is the standard deviation of the investment return, σ_i, using the formulas in Kreps' appendix. Again, I won't show those here, but the resulting value is $\sigma_i = 5.69\%$.

Next, using the same logic as before, we want to ensure that the variance of our combined reinsurance transaction and put-protected investment of the premium and invested capital is less than or equal to the variance of our investment return had we instead invested that capital directly in the risky asset. In this case, the solution for A is far messier that it seems like it should be, but using the formulas shown above we can solve for $A = 32,522,638$.

Finally, the risk load for the put option strategy under the variance constraint is found using the same formula as before but with this revised value for A:

$$\begin{aligned} R &= A\left[\frac{(1+r)(1+y)-(1+i)}{1+i}\right] + \mu_L\left[\frac{1+r}{1+i} - \frac{1}{1+r_f}\right] \\ &= (32,522,638)\left[\frac{(1.0318)(1.053)-(1.0784)}{1.0784}\right] + 1,000,000\left[\frac{1.0318}{1.0784} - \frac{1}{1.036}\right] \\ &= 235,217 \end{aligned}$$

And as before, we select the higher of the two risk loads under the two constraints, which in this case is the risk load under the variance constraint, or $R = 235,217$.

Review of Procedure

Don't be confused by all of the formulas, since many of them are unlikely to be tested. Instead, focus primarily on the process.

First use the swap strategy and calculate the risk load under both the safety and the variance constraints. Whichever leads to the biggest value for the invested capital, and hence the biggest risk load, is the one to use for this strategy. Then do the same for the put option strategy, selecting the higher of the two risk loads that result from the two constraints. Finally, the risk load to charge is the *lower* of the swap or put option strategy results.

Minimum Rate on Line for High Excess Layer

Suppose you used the put option strategy in the case where the reinsurance is for a very high excess layer. In this case, with a large limit and a very small probability of loss, the safety constraint will be the dominant one and the safety level, s, can reasonably be assumed to be equal to the limit. Setting μ_L to zero and using the put option/safety constraint formula for the risk load we have:

$$R = s\left[\frac{(1+y)(1+r)-(1+i)}{(1+r_f)(1+y)}\right]$$

If we take s to be the policy limit, then the Rate on Line (ROL) is given by the ratio of R to s. The minimum rate on line for $\mu_L = 0$ is given by:

$$\text{ROL} = \left[\frac{(1+y)(1+r)-(1+i)}{(1+r_f)(1+y)}\right]$$

That's for the put option strategy. For the swap strategy,

$$\text{ROL} = \frac{y-r_f}{(1+r_f)(1+y)}$$

For realistic values, the put option strategy will usually produce the lower of the two risk loads. And recall that these formulas are for the special case where the layer gets high enough so that the probability of loss approaches zero.

Single Payment at Arbitrary Time

The formulas above all assumed that the losses are paid at a single point in time, one year from the pricing date. If the payment occurs at any arbitrary point in time, t, but is still a single payment, then the formulas are trivial to modify. For each of the investment rate, the risky asset yield and the risk free rate we simply replace $(1+r^*)$ with $(1+r^*)^t$.

The trickier calculations when the payment is not at time $t = 1$ are those for the yield standard deviation (σ_y), the mean of the put-protected investment return (i) and the standard deviation of the put-protected investment return (σ_i). In particular, we need to convert the yield mean and standard deviation into the lognormal parameters for the distribution of $(1+y)$, then use μt and $\sigma\sqrt{t}$ as those distribution parameters and convert back to the mean and standard deviation variables y and σ_y.

Pooling Across Transactions

Keep in mind that Kreps has developed a risk load formula that ignores existing reinsurance risks in the portfolio. However, we can consider what would happen to the risk load if we were to simultaneously take on two reinsurance contracts that are imperfectly correlated and compare it to the sum of the risk loads for each.

Intuitively, what happens is the distribution for the reinsurance claims becomes less skewed, which lowers the safety level and should result in a combined risk load that is smaller than the sum of the stand-alone risk loads.

Numerical Example

As an example, what if we wrote two policies just like the one we've been using and assumed they are uncorrelated. In this case, the mean claim costs would be $\mu_L = \$2,000,000$ and the standard deviation would be $\sigma_L = \$2,828,427$. Then, repeating all of the same calculations used earlier, we get the results shown in Table 1.

TABLE 1. Risk Loads with Two Independent Risks

Strategy and Constraint	Risk Load
Swap - Variance Constraint	552,529
Swap - Safety Constraint	427,839
Swap Risk Load	552,529
Option - Variance Constraint	186,995
Option - Safety Constraint	331,146
Option Risk Load	331,146
Selected Risk Load	**331,146**

Notice that the risk load when two independent risks are pooled is only about 70% of the sum of the individual risk loads. But of course, it is larger than the risk load needed for just one transaction, which is the point made in the discussion of risk pooling and risk sharing in the BKM readings.

What Makes the Kreps Method *Investment Equivalent*?

Writing reinsurance will never result in distributions of returns on capital that are identical to the distribution of returns on invested assets. Investment returns are strictly limited to having worst-case returns of -100%, whereas any company writing a reinsurance policy could face claims that exceed the premium charged and the capital committed. Unless this is the only policy they write and/or the only financial resources available to them, their return on invested capital distribution can certainly be less than -100% of their initial invested capital.

In addition, an insurer writing multiple contracts would certainly not want to actually implement either the swap or put option strategy separately for each transaction.

However, what if a legal entity were set up specifically to write a single transaction or a pool of separate transactions executed simultaneously, as in the case of catastrophe bonds and sidecars (see the Cummins reading)? In this case, the Kreps approach might be both practical to implement and useful for establishing the risk load. The latter point is important because recall that Kreps pointed out that the risk loads he derives are conservative because they don't reflect the rest of the company's exposures. This would not be the case if this were indeed the only transaction.

Further, while in the numerical examples we have taken the claim distribution to be given and solved for the assets that had to be invested under different constraints, we could just as well have taken the invested capital as an input and derived the constraints on the loss distribution. This information could be used to establish underwriting standards, for instance.

Practice Questions

Question 1. Briefly describe the two financial strategies used by Kreps to develop his various risk loads.

Solution. Kreps' goal is to use a financial strategy, along with writing a reinsurance contract, that ensures that the total funds invested are not exposed to the risk of earning less than the risk free rate of return. His two methods are:

1. Swap Risk-Free Investment for Risky Investment — Invest the funds, comprising the premiums charged (net of expenses) plus the required capital investment, in risk free investments rather than in risky assets.

2. Purchase Put Options — Invest in risky assets but also purchase a put option with a strike price equal to the value of the funds grown at the risk free rate. This eliminates the downside risk from the investment in risky assets.

Question 2. To ensure that writing a reinsurance policy and simultaneously following one of the two financial strategies described in the previous question can be compared to a direct investment in risky assets, Kreps requires the expected rate of return in the two alternatives be equal. But he also applies two additional constraints to ensure their risk is comparable as well. What are these two constraints?

Solution.

a. Safety Constraint — The safety constraint simply requires that the funds available to pay claims at the end of the year are at least equal to the specified loss safety level.

b. Investment Variance Constraint — The variance constraint simply requires that the IRR of the combined reinsurance and investment strategy be no more volatile than a direct investment in the risky assets.

Question 3. You are considering writing a reinsurance contract on a risk with lognormally distributed claim costs, for which the mean is $\mu_L = 3,000,000$ and the standard deviation is $\sigma_L = 3,464,102$. This results in lognormal distribution parameters of $\mu = 14.490$ and $\sigma = 0.9205$.

Assume that all claims are paid at the end of one year and that the risk free rate is 3.6%. If writing this contract requires a capital investment of $A = 13,025,196$, what is the expected internal rate of return on the invested capital if the insurer were to charge a premium of $P = 3,000,000$ and follow Kreps' swap strategy?

Solution. Under the swap strategy, the insurer will commit an amount A at the beginning of the year, invest this amount and the charged premium in the risk-free investment and then pay out claims with an expected value of 3,000,000 at the end of the year. The internal rate of return on this is simply the rate that solves the following equation:

$$(1 + E(\text{IRR}))A = (1 + r_f)(P + A) - \mu_L$$

Just in case that isn't blatantly obvious, just note that the IRR is simply the rate that sets the initial investment equal to the present value (when discounted at the IRR) of all subsequent cash flows. Therefore, it might be easier to see this as:

$$A = \frac{(1 + r_f)(P + A) - \mu_L}{(1 + E(\text{IRR}))}$$

In any case, it is easy to solve this equation for the expected IRR since we were given all of the other values:

$$\begin{aligned} E(\text{IRR}) &= \frac{(1 + r_f)(P + A) - \mu_L}{A} - 1 \\ &= \frac{(1.036)(3,000,000 + 13,025,196) - 3,000,000}{13,025,196} - 1 \\ &= 4.429\% \end{aligned}$$

Question 4. If you wanted to ensure that the expected IRR from the previous question were equal to the yield on a portfolio of risky assets, $y = 5.3\%$, what would the risk load have to be equal to?

Solution. This is really just a matter of setting the last equation of the previous solution equal to 5.3% and solving for the premium P.

$$\begin{aligned} E(\text{IRR}) &= \frac{(1 + r_f)(P + A) - \mu_L}{A} - 1 \\ 5.3\% &= \frac{(1.036)(P + 13,025,196) - 3,000,000}{13,025,196} - 1 \\ P &= 3,109,487 \end{aligned}$$

The question asked for the risk load though, so you need to solve for R in the following:

$$P = R + \frac{\mu_L}{1 + r_f} \Rightarrow R = 213,734$$

That's the logic of part of the Kreps approach, though notice that so far I have told you what the value of A was. In practice, we need to calculate that by applying his safety and variance constraints, which we will do in the next problems.

Also notice that we could have solved for the risk load directly using the formula provided in the paper for a known quantity A:

$$\begin{aligned} R &= \frac{y - r_f}{1 + r_f} A \\ &= \frac{.053 - .036}{1.036}(13,025,196) \\ &= 213,734 \end{aligned}$$

Question 5. Using the same facts as above, apply Kreps' safety constraint with a safety level set equal to the 99th percentile of the claim cost distribution to jointly determine the amount of initial capital to invest, A, and the risk load, R, assuming the use of the swap financial strategy. Recall that for a Standard Normal distribution the 99th percentile is 2.33 standard deviations above the mean, $z = N^{-1}(.99) = 2.33$

Solution. Luckily we don't have to derive the lognormal distribution parameters from the mean and standard deviation as those were given to us in the question as $\mu = 14.490$ and $\sigma = 0.9205$. From this, we can see that the safety level is:

$$s = e^{\mu + z\sigma} = e^{14.49 + (2.33)(.9205)} = 16,715,531$$

Now, we can apply the safety constraint, which requires that the premium charged and the capital invested grow at the risk-free rate to be at least equal to the safety level, s.

$$(1 + r_f)(P + A) \geq s$$

We also know from swap strategy that we want the IRR to equal the yield on the risky asset portfolio. Starting with the formulas used in the previous question I will derive a formula for the quantity $(1 + r_f)(P + A)$ that can be plugged into the above inequality and simplified:

$$\begin{aligned} E(\text{IRR}) &= \frac{(1 + r_f)(P + A) - \mu_L}{A} - 1 \\ y &= \frac{(1 + r_f)(P + A) - \mu_L}{A} - 1 \\ (1 + y)A + \mu_L &= (1 + r_f)(P + A) \end{aligned}$$

Plugging this into the inequality and simplifying we get:

$$\begin{aligned}(1+y)A+\mu_L &\geq s\\ A &\geq \frac{s-\mu_L}{1+y}\\ &\geq \frac{16,715,531-3,000,000}{1.053}\\ &\geq 13,025,196\end{aligned}$$

From this value for A at the equality, the risk load is found using that same formula as before:

$$\begin{aligned}R &= \frac{y-r_f}{1+r_f}A\\ &= \frac{.053-.036}{1.036}(13,025,196)\\ &= 213,734\end{aligned}$$

Question 6. Assume that the standard deviation of the risky investment yield were 8.4%. If the variance constraint were used in the previous question, what would be the value for the invested capital, A?

Solution. Under this constraint, we want the dollar standard deviation of the combination of the reinsurance contract and the risk-free investment to be less than or equal to the dollar standard deviation of the same invested capital invested in the risky asset.

$$\begin{aligned}A\sigma_y &\geq \sigma_L\\ A &\geq \frac{\sigma_L}{\sigma_y}\\ &\geq \frac{3,464,102}{.084}\\ &= 41,239,309\end{aligned}$$

Question 7. Continuing to use the same facts as in the previous question, apply the variance constraint and the swap financial strategy under the assumption that instead of being paid in one year the reinsurance claims are paid in three years (all at once). Determine the value for A and then calculate the risk load in this case.

Solution. Start with the value for A, which is easy but perhaps not very clear in the reading. From the previous question we know what the formula is and we know that the mean claim amount does not change. But what does change is that the standard deviation of the three-year yield is not the same as the one-year yield, so calculations are needed.

We assume that the risky investment yield (one plus the yield, actually) is lognormally distributed and therefore we have to first calculate the lognormal distribution parameters using the mean and variance provided.

The formulas we need are:

$$\sigma = \sqrt{\ln\left(1 + \left(\frac{\sigma_y}{1+y}\right)^2\right)}$$

$$\mu = \ln(1+y) - \frac{\sigma^2}{2}$$

Using these, we can solve for $\mu = 0.0485$ and $\sigma = 0.0796$. These are the one-year parameters, which can be adjusted to reflect the three-year values using $\mu t = 0.1454$ and $\sigma\sqrt{t} = 0.1380$. And then, finally, we use those to calculate the mean and standard deviation:

$$\text{Mean} = 1 + y = e^{\mu + \frac{\sigma^2}{2}} = 1.1676$$

$$\sigma_y = \sqrt{e^{\sigma^2} - 1}(1+y) = 16.18\%$$

And now, we can recalculate the value for A under the variance constraint:

$$A = \frac{\sigma_L}{\sigma_y} = \frac{3,464,102}{.1618} = 21,404,919$$

Now, we just need to plug in for the formula for the risk load under the swap strategy:

$$R = \frac{y - r_f}{1 + r_f} A$$

The only tricky part is to recognize that for all of the interest related variables (y, r_f, etc.) we need to reflect their three-year values. In this case, we use $y = 16.76\%$ from the calculations above and $r_f = 1.036^3 - 1 = 11.19\%$:

$$\begin{aligned} R &= \frac{y - r_f}{1 + r_f} A \\ &= \frac{.1676 - .1119}{1.1119}(21,404,919) \\ &= 1,071,102 \quad \text{without rounding} \\ &= 1,072,267 \quad \text{using the rounded values shown} \end{aligned}$$

Question 8. Suppose you wanted to use Kreps' put option strategy along with the risky asset used in the previous questions, with an annual mean yield and standard deviation of yield equal to $y = 5.3\%$ and $\sigma_y = 8.4\%$. As part of this strategy the premium and some of the invested capital will be invested in the risky asset and some of the invested capital will be used to purchase a put option. Explain why the expected investment return is not equal to $y = 5.3\%$ and determine whether the expected investment return is higher or lower than y.

Solution. Under the put option strategy the end-of-period invested assets will equal the maximum of the beginning assets multiplied by the actual investment performance of the risky asset portfolio if the put option is not exercised or the beginning assets grown at the risk free rate in the case where the put option is exercised. Since this eliminates the downside risk in the risky asset yield, it's mean has to be higher than the expected risky asset yield, y.

For instance, using the values for the key assumptions we've been using so far and assuming the reinsurance cash flows occur in one year, the expected investment return under the put option strategy would be $i = 7.84\%$.

The calculations to get this are shown in Kreps' appendix but do not appear to be needed for the exam.

Question 9. Suppose you were going to use Kreps' put option strategy in the case where the risk-free rate is 3.6%, the risky investment yield is 5.4% and its standard deviation is 8.4%. What would be the cost of the put option per dollar of invested assets, where the invested assets are denoted F?

Solution. Under this strategy, the put option has a strike price equal to the forward value of the invested assets. As a consequence, in this special case the formula for the value of the put option is relatively simple:

$$\text{Option Cost} = F\left[\Phi\left(\frac{\sigma\sqrt{t}}{2}\right) - \Phi\left(\frac{-\sigma\sqrt{t}}{2}\right)\right]$$

As a rate, we just divide by the amount invested to get:

$$r = \Phi\left(\frac{\sigma\sqrt{t}}{2}\right) - \Phi\left(\frac{-\sigma\sqrt{t}}{2}\right)$$

But there's a complication associated with calculating this from the inputs given. It is tempting to use 8.4% as the value of the σ variable above. However, that quantity given is the yield standard deviation, σ_y, and not the lognormal distribution parameter we need. We saw the number we need in an earlier problem:

$$\sigma = \sqrt{\ln\left(1 + \left(\frac{\sigma_y}{1+y}\right)^2\right)} = 0.0796$$

From this, the put option cost as a rate is:

$$\begin{aligned} r &= \Phi\left(\frac{\sigma\sqrt{t}}{2}\right) - \Phi\left(\frac{-\sigma\sqrt{t}}{2}\right) \\ &= \Phi\left(\frac{0.0796}{2}\right) - \Phi\left(\frac{-0.0796}{2}\right) \\ &= 0.0318 \end{aligned}$$

Notice that there are two approximations to keep in mind, which might be useful:

$$\sigma \approx \frac{\sigma_y}{1+y} = \frac{.084}{1.054} = .0797$$

$$r \approx .4\sigma\sqrt{t} = .4(.0796) = .03185$$

These might be able to save you some time on the exam.

Question 10. In the earlier questions involving the swap strategy we used the objective that we wanted the IRR of the combined reinsurance transaction and financial strategy to be equal to the yield on the risky asset, y, and showed that this could be intuitively written (before simplifying) as:

$$(1 + y)A = (1 + r_f)(P + A) - \mu_L$$

because $(P + A)$ reflects the total amount invested in the risk-free asset and so the right-hand side shows the final cash flows available to us (if positive) at the end of the year and if this were discounted at the risky asset yield of y we would have our initial investment, A.

Explain which of the values in the previous formula would be changed if instead we used the put option strategy.

Solution. There are really just two changes that have to be made.

The first is to recognize that we are no longer investing the funds $(P + A)$ because we need to use some of our funds to purchase the put option. Therefore, the invested funds are given by the quantity:

$$\begin{aligned} F &= P + A - rF \\ &= \frac{P + A}{1 + r} \end{aligned}$$

The second change is that we are no longer investing these funds at the risk-free rate, so they will grow to the uncertain return on the put-protected risky investment, with a mean value of i.

Simply replacing these two components gives us the revised relationship:

$$(1 + y)A = (1 + i)F - \mu_L$$

$$(1 + y)A = (1 + i)\frac{P + A}{1 + r} - \mu_L$$

And since $P = \frac{\mu_L}{(1+r_f)} + R$ we can plug this quantity in to this equation and rewrite in terms of the risk load, R as:

$$R = A\left[\frac{(1 + r)(1 + y) - (1 + i)}{1 + i}\right] + \mu_L\left[\frac{1 + r}{1 + i} - \frac{1}{1 + r_f}\right]$$

Question 11. In the previous question we saw that in order to ensure that the expected return (IRR) on contributed capital (A) under the put option strategy is the same as the yield that would be achieved if the contributed capital were simply invested in the risky asset, the relationship between the contributed capital and the amount actually invested in the risky asset (after paying for the put option) is:

$$(1 + y)A = (1 + i)F - \mu_L$$

Use this, along with the safety constraint, to derive the formula for the amount of capital that has to be contributed.

Solution. The safety constraint requires that whatever funds we invest, F, grow to be at least as big as the safety level, s. In the put option case, the funds will grow to *at least* $(1 + r_f)F$ because we have a put that ensures this is the case. So the safety constraint just says that this worst case must be equal greater than or equal to s, or:

$$(1 + r_f)F \geq s$$

At the equality,

$$F = \frac{s}{(1 + r_f)}$$

This can be plugged into the first equation to get:

$$(1 + y)A = (1 + i)F - \mu_L$$

$$(1 + y)A = (1 + i)\left[\frac{s}{1 + r_f}\right] - \mu_L$$

$$A = \frac{1}{1 + y}\left[\frac{1 + i}{1 + r_f}s - \mu_L\right]$$

Question 12. Write the variance constraint in the case of the put option strategy and show why in order to write a formula for the required invested capital, A, we must solve a complicated quadratic equation. Assume the investment performance and the reinsurance claims are independent.

Solution. Begin with the variance constraint, which states that the variance of the combination of the reinsurance transaction and the put-protected investment has to be less than or equal to the variance had the invested capital simply been invested in the risky asset. This constraint can be written as:

$$(A\sigma_y)^2 \geq (F\sigma_i)^2 + (\sigma_L)^2$$

where, A is the invested capital to support the reinsurance transaction and F is the total amount of funds invested in the risky assets when reinsurance is combined with the put option strategy. With respect to the standard deviation terms, σ_y is the standard deviation of the risky asset yield, σ_i is the standard deviation of the rate of return on the put-protected investment in the risky asset (with the downside risk eliminated) and σ_L is the standard deviation of the reinsurance claims.

Note, by the way, that σ_i is rather complicated to calculate as it involves a random variable that protected on the downside to be no less than the risk-free rate.

Now, to get the formula for A under this constraint and the requirement that the IRR on the combination of reinsurance and the put strategy must equal the risky asset yield, we plug the formula for F in terms of the risky asset yield:

$$F = \frac{(1+y)A + \mu_L}{(1+i)}$$

Into the variance constraint formula at the equality:

$$A^2\sigma_y^2 = \left[\frac{(1+y)A + \mu_L}{(1+i)}\right]^2 \sigma_i^2 + \sigma_L^2$$

But notice that we have to square the term in the brackets, which gives us a complicated term involving A^2 on the right-hand side, as well as one on the left-hand side terms involving A as well. This will require the quadratic formula to solve, with the resulting solution being quite a mess:

$$A = \frac{b + \sqrt{b^2 + ac}}{a}$$

where

$$a = \sigma_y^2(1+i)^2 - \sigma_i^2(1+y)^2$$

$$b = \mu_L(1+y)\sigma_i^2$$

$$c = \mu_L^2\sigma_i^2 + \sigma_L^2(1+i)^2$$

Question 13. You have calculated risk loads, shown in the table below, using each of the four combinations of financial strategies and risk constraints for a proposed reinsurance transaction with expected claim costs of \$2 million and a standard deviation of claim costs of \$2.8 million. If the risk free rate is 3.6%, all claims are paid in two years and company expenses are 15% of premium, what would be the premium charged for this transaction following the process proposed by Kreps?

Strategy and Constraint	Risk Load
Swap - Variance Constraint	746,965
Swap - Safety Constraint	334,998
Option - Variance Constraint	123,963
Option - Safety Constraint	445,267

Solution. Kreps proposed, for each financial strategy, calculating both the risk loads subject to both the variance and safety constraints and selecting the larger of the two such that both constraints are satisfied. Then, choose the risk load under the financial strategy that produces the lowest risk load so that the premium is competitive. This is done in the following table:

Using this risk load, the pure premium is:

$$P = R + \frac{\mu_L}{(1+r_f)^2} = 445,267 + 2,000,000/1.036^2 = 2,308,686$$

Strategy and Constraint	Risk Load
Swap - Variance Constraint	746,965
Swap - Safety Constraint	334,998
Swap Risk Load	**746,965**
Option - Variance Constraint	123,963
Option - Safety Constraint	445,267
Option Risk Load	**445,267**
Selected Risk Load	**445,267**

The final premium simply reflects a gross up for expenses, so:

$$P' = 2,308,686/(1 - .15) = 2,716,101$$

Question 14. You are considering a high layer excess of loss reinsurance contract where the loss probability is small enough so that the distribution can be assumed to be binomial and the mean loss approaches zero? Use the put option strategy and the safety constraint to determine the minimum rate on line. Assume the safety level is set equal to the full policy limit, the risk-free rate is 3.6%, the premium and capital funds will be invested in risky assets with an expected yield of 5.3% and a yield standard deviation of 8.4%, the put is priced at a rate of 3.18% and the expected put-protected return is 7.84%.

Solution. Kreps' formulas for the risk load under the put option strategy and the safety constraint is:

$$R = \frac{1}{(1+r_f)(1+y)}\left[s((1+y)(1+r) - (1+i)) - \mu_L(y - r_f)\right]$$

But with $\mu_L \approx 0$ we can simplify this to the formula for the pure premium:

$$P = s\left[\frac{(1+y)(1+r) - (1+i)}{(1+r_f)(1+y)}\right]$$

Taking s equal to the policy limit and writing the rate on line (ROL) as the premium relative to the limit this becomes:

$$\begin{aligned}\text{ROL} &= \left[\frac{(1+y)(1+r) - (1+i)}{(1+r_f)(1+y)}\right]\\ &= \left[\frac{(1.053)(1.0318) - (1.0784)}{(1.036)(1.053)}\right]\\ &= .74\%\end{aligned}$$

Question 15. Using the results from the previous question, show how the formula for the ROL can be easily modified to reflect the ROL for the swap strategy under the same safety constraint and the same assumptions.

Solution. Notice that the swap strategy would result in an expected investment return $i = r_f$ and a put cost of $r = 0$. Simply plugging these values into the ROL formula gives:

$$
\begin{aligned}
\text{ROL} &= \left[\frac{(1+y)(1+r)-(1+i)}{(1+r_f)(1+y)}\right] \\
&= \left[\frac{(1+y)-(1+r_f)}{(1+r_f)(1+y)}\right] \\
&= \left[\frac{(y-r_f)}{(1+r_f)(1+y)}\right] \\
&= \left[\frac{.053-.036}{(1.036)(1.053)}\right] \\
&= 1.56\%
\end{aligned}
$$

Selected Old Exam Questions for Part 7

The following questions relevant for this section appeared on the Old CAS Exam 9 between 2002 and 2010 and on the New CAS Exam 9 since 2011.

Ferrari	**Roth**	**McClenahan**	**Robbin**	**Mango**	**Kreps**
2002 Q27	2002 Q24	2002 Q5	2002 Q21	2004 Q34	2004 Q37
2003 Q41	2002 Q4	2003 Q16	2003 Q42	2005 Q19	2008 Q16
2005 Q20	2003 Q19	2003 Q40	2003 Q44	2006 Q19	2009 Q10
2006 Q18	2004 Q11	2004 Q10	2003 Q45	2006 Q20	2010 Q8
2007 Q13	2004 Q8	2005 Q22	2004 Q12	2007 Q20	2012 Q15
2007 Q17	2005 Q12	2006 Q22	2004 Q13	2007 Q23	2013 Q22
2008 Q17	2005 Q14	2008 Q19	2004 Q32	2008 Q11	2013 Q23
2009 Q8	2006 Q13	2012 Q16	2004 Q33	2009 Q11	2014 Q23
2009 Q9	2007 Q14	2014 Q21	2004 Q36	2009 Q12	2015 Q26
2010 Q7	2008 Q13	2015 Q22	2005 Q13	2010 Q14	
2010 Q9	2014 Q24		2006 Q12	2011 Q17	
2011 Q15	2015 Q24		2006 Q17	2012 Q13	
2012 Q12	2015 Q25		2007 Q16	2013 Q21	
2013 Q24			2007 Q18	2014 Q22a	
2013 Q25			2007 Q19	2015 Q27	
2015 Q21			2008 Q12		
			2008 Q15		
			2009 Q14		
			2010 Q10		
			2010 Q12		
			2010 Q15		
			2011 Q18		
			2011 Q20		
			2011 Q21		
			2012 Q18		
			2013 Q17		
			2013 Q18		
			2013 Q19		
			2014 Q19		
			2014 Q20		
			2015 Q23		

For some of these questions I have provided the text of the question and an explanatory solution. These were selected either because they are representative of the questions you are likely to be asked on future exams or because they contain an element that is particularly worthwhile to point out. For the other questions, the CAS solutions should be sufficient to confirm whether your answer is correct.

Given the difficulty of some of the past questions related to the Robbin and Feldblum papers, I will cover those first before proceeding to some of the other questions from this section of the syllabus.

Important Note: The solutions shown here are intentionally detailed. They contain thorough explanations of the concepts and formulas used to reinforce the main points from the readings and provide an additional teaching opportunity. **Actual exam responses should be much more concise than what is shown here, along the lines of what you will see in the solutions that the CAS releases.**

2010 Exam Question 7

A monoline insurance company is operating under the following conditions:

Expected Loss	500
Ratio of Undiscounted Reserves to Surplus	2.0
Required Rate of Return	15%
Investment Yield	5%
Risk Free Rate	3%

Additionally, you are told all premium is collected on Jan 1, 2011; all losses are paid on Jan 1, 2012; there are no other expenses or taxes; the risk is independent of all other risks and the risk has a one-year policy term effective Jan 2, 2011.

Note: The question was not clear, but for the purposes of my answer I am assuming that the new risk is just like the old risks, so that it has an expected losses of $500.

Calculate the minimum amount of premium to be charged to the new risk to meet the required return.

There are multiple ways to interpret this question because no specific method was identified. The simplest is to recognize that the fact that the policy term is one year means that there is no difference between policy year, accident year and calendar year. This means we can use Ferrari's relationship for *calendar year* data:

$$\frac{T}{S} = \frac{I}{A}\left(1 + \frac{R}{S}\right) + \frac{U}{P}\frac{P}{S}$$

If we want the return on surplus to be 15%, then we have to solve for U in the above equation:

$$15\% = 5\%\,(1 + 2) + \frac{U}{500/2}$$

Notice that I assumed that at inception there was a total of $500 of expected losses posted as a reserve and held throughout the year. Depending on when the premium is earned this reserve might initially be an unearned premium reserve and then eventually move to a loss reserve, but either way reserves will be $500.

Also note that I replaced the underwriting profit as a ratio to premium and the premium to surplus ratios so that I could avoid having an extra term with the unknown premium, which would just complicate the algebra. That could be avoided because we know the dollar amount of surplus already from the reserve and the reserve to surplus ratio.

Solving the algebra gives us $U = 0$. This is obvious because we know that our investors earn 5% investment return on their surplus and 5% on the reserves which are twice the surplus. That means that they earn 15% on their surplus, which is the target return without yet considering any underwriting profit.

Finally, to get the premium we would have to add the underwriting profit (in dollars in this case) to the expected loss to get $P = 500$.

Notice that I didn't use discounted expected loss since the return calculation above already accounted for the investment income on the reserves until losses are paid.

As an alternative, you could have used Robbin's Calendar Year ROE method. The Robbin formula before converting to his notation is:

$$ROE = \frac{\text{Underwriting Profit Provision in Dollars} + \text{Investment Income} - \text{Taxes}}{\text{Surplus}}$$

All we have to do is determine the dollar amount of the investment income and ignore the taxes. We know that the investment income consists of the portfolio yield on both the policyholder supplied funds (the total reserves of 500 in this case) and the surplus.

So the above equation becomes:

$$15\% = \frac{U_{\$} + 5\%(500 + 250)}{250} \Rightarrow U = 0$$

2010 Exam Question 10

An insurer is considering a new risk with the following characteristics:

Percentage of Premium Collected on Jan 1, 2011	100%
Present Value of Federal Income Tax at Dec 31, 2011	4
Risk-free Rate per Quarter	2%
Risk-Adjusted Rate per Quarter	1.50%

Losses and expense are expected to be paid according to the following chart:

Payment Date	Expected Loss	Expected Expenses
1/1/2011		10
3/31/2011	10	5
6/30/2011	15	5
9/30/2011	20	5
12/31/2011	15	5
3/31/2012	5	
6/30/2012	5	

Use the risk-adjusted discounted cash flow method to calculate the full premium for this risk.

This method is quite simple once you determine the proper risk-adjusted discount rate, which was specified in the question. We calculate the fair premium as being equal to the present value of all loss and expense cash flows, including *taxes paid on the present value of underwriting income and taxes paid on the investment income on surplus.*

But notice three wrinkles.

First, the treatment of taxes is sometimes a bit more complex than it seems and the exam questions don't always provide the clearest information. Here we are told the present value of federal income taxes, as of the end of the period, so it has to be assumed that this includes all of the taxes intended to be included, which are the taxes on underwriting income and the tax on the investment income on surplus.

Secondly, note that in the paper Robbin's numerical example discounted all cash flows to the end of the first year. It is never very clear on the exam problems how this is intended to be done because if you calculate the value of the premium at the end of the year you still need to discount that back one period to know what the premium is today, however the numbers differ slightly because of minor differences in how the loss cash flows in the first year are discounted. I will ignore this issue and calculate all present values to today.

And third, in the paper Robbin actually uses risk-free rates for all cash flows other than the loss cash flows, but this is just a simplification in the paper. It does not have to be done this way, assuming one could justify an appropriate risk-adjusted rate for those other cash flows. Nonetheless, it seems appropriate to do that here as well.

Now, the basic formula we are going to use, reflecting that all of the premium is paid up front, is:

$$P = PV(L; i_r) + PV(FX; i_f) + PV(FIT; i_f)$$

The PV(L) is found from the stream of payments, discounting at the risk-adjusted rate of 1.5% per quarter. The PV(FX) is done similarly, but using the risk free rate of 2.0% per quarter.

Payment Date	Nominal Loss	Disc Factor	PV(Loss)	Nominal Exp	Disc Factor	PV(Exp)
1/1/2011	0.00	1.000	0.00	10.00	1.0000	10.00
3/31/2011	10.00	0.985	9.85	5.00	0.9804	4.90
6/30/2011	15.00	0.971	14.56	5.00	0.9612	4.81
9/30/2011	20.00	0.956	19.13	5.00	0.9423	4.71
12/31/2011	15.00	0.942	14.13	5.00	0.9238	4.62
3/31/2012	5.00	0.928	4.64	0.00	0.9057	0.00
6/30/2012	5.00	0.915	4.57	0.00	0.8880	0.00
Present Value			66.89			29.04

Finally, the taxes were given as $4 at the end of the year and we can discount those one year to today, again at the risk-free rate. Using the 12/31/2011 discount factor used for the expenses of 0.9238, the present value of the taxes is 3.70.

PV(Loss)	66.89
PV(Expense)	29.04
PV(FIT)	3.70
Premium	**99.62**

2010 Exam Question 12

An insurance company's most recent financial statement reflects the following:

Direct Earned Premium	100,000.00
Incurred Losses	55,000.00
Fixed Expenses	10,000.00
Variable Expenses as Percent of Premium	25.00%
Policyholder Supplied Funds	60,000.00
Premium to Surplus Ratio	3.20
Premium to Equity Ratio	2.50
Pre-Tax Investment Yield	8.00%
income Tax Rate	34.00%

a. Use the calendar year return on equity procedure to calculate the return on equity.

The CY ROE method is simply the ratio of after-tax income divided by GAAP equity. But in numerical problems like this we often need to calculate various components. What we are ultimately trying to get to is:

$$\begin{aligned} ROE &= \frac{\text{Underwriting Profit Provision in Dollars} + \text{Investment Income} - \text{Taxes}}{\text{Equity}} \\ &= \frac{U \cdot P + II - FIT}{EQ} \\ &= \frac{U \cdot P \cdot (1-t) + II(1-t)}{EQ} \end{aligned}$$

Notice in the last line I replaced the underwriting profit and the investment income with their after-tax amounts. This helps to keep the calculations a bit simpler.

For the investment income, we are going to want to include both the surplus, which we calculate as $100,000/3.2 = 31,250$ and the policyholder supplied funds of $60,000$.

With this, the two after-tax income components are easy to calculate as follows, with all amounts in dollars to make it simple:

$$\begin{aligned} UP(1-t) &= [100,000 - 55,000 - [10,000 + .25(100,000)]](1-.34) = 6,600 \\ II(1-t) &= [8\%(31,250 + 60,000)](1-.34) = 4,818 \\ EQ &= 100,000/2.5 = 40,000 \end{aligned}$$

Finally, CY ROE $= (6,600 + 4,818)/40,000 = 28.55\%$.

b. Discuss two advantages of the calendar year return on equity procedure.

The CY ROE method is appealing because most of the inputs are readily obtained from reported financial statements and exhibits and it produces a metric that is similar to the GAAP ROE commonly used to measure profitability in other industries.

2010 Exam Question 15

An insurance company uses the present value offset method to set insurance rates. The reference line has a traditional underwriting profit provision of 4% and losses are paid at the end of the year according to the pattern shown below.

The insurance company is reviewing its rates for a separate line of business and notes the following:

Underwriting Profit Provision for Reviewed Line of Business	2%
Discount Rate	5%
Variable Expense Ratio	20%
Fixed Expenses	200.00
Undiscounted Losses for the Reviewed Line of Business	1,000.00

Losses for the reference line and the reviewed line are paid at the end of the year according to the following pattern:

Payment Patterns		
Year	Reference LOB	Reviewed LOB
1	40%	20%
2	30%	30%
3	30%	40%
4	0%	10%

a. Use the present value offset method to calculate the permissible loss ratio for the reviewed line of business.

This is a bit of a strange question and seems to contain inconsistent information. Let me first give you the answer as they expected it and then show you what I think is wrong with the information given. Basically, the two underwriting provisions given are inconsistent with the rest of the information.

They appear to have wanted you to use a formula from the Robbin paper that allows you to solve for the permissible loss ratio (PLR). He showed a formula for the reviewed line's underwriting profit provision (U) as a function of the reference line's underwriting profit provision (U^0), the permissible loss ratio (PLR) and the differences in the two lines' loss discount factors ($PV(X^0) - PV(X)$).

$$U = U^0 - (PLR)[PV(X^0) - PV(X)]$$

To use that formula we need to know the present value discount factors for the reference payment pattern (X^0) and the payment pattern for the reviewed line (X). These are calculated as follows:

		Reference Line (X^0)		Reviewed Line (X)	
Year	$(1+r)^{-i}$	Payout	Payout * Disc. Factor	Payout	Payout * Disc. Factor
1	0.95238	40%	0.3810	20%	0.1905
2	0.90703	30%	0.2721	30%	0.2721
3	0.86384	30%	0.2592	40%	0.3455
4	0.82270	0%	0.0000	10%	0.0823
			0.9122		**0.8904**

Then plugging into the formula given we have:

$$U = U^0 - (PLR)[PV(X^0) - PV(X)]$$

$$2\% = 4\% - (PLR)[.9122 - .8904]$$

Solving for PLR we get $PLR = 91.65\%$

But notice that to use that formula I needed to know the "correct" underwriting profit provision for the reviewed line that is consistent with the underwriting profit provision for the reference line. The two are not consistent. To be fair, they didn't say that the 2% provision in the reviewed line is the right one. Perhaps that was meant to represent the current one that you are trying to update.

But then it isn't clear if they want the permissible loss ratio used in the current pricing or the loss ratio implied by the correct underwriting profit provision. Let's move on to the next part and come back to this inconsistency.

b. Use the present value offset method to calculate the premium for the reviewed line of business.

This method assumes that the underwriting profit provision (UPP) for the reference line is "correct" and that it incorporates already the effect of discounting of losses. The method then simply adjusts the UPP for other lines of business based upon their relative speed of loss payment, so that slower paying lines with more discounting have lower UPPs.

The formula you need to use is simple. The premium to charge should be calculated directly as follows:

$$P = L[PV(X)] + FX + VX(P) + U^0(P) + L[1 - PV(X^0)]$$

The first term reflects present value losses for the line being reviewed. The last term reflects the amount of discounting of those losses that is implicitly buried in the traditional underwriting profit provision, U^0. It is being added back because we assume it is already accounted for and since we are explicitly discounting losses at the right pattern, we don't need to include the implicit discount amount too.

Using this simple formula,

$$\begin{aligned} P &= L[PV(X)] + FX + VX(P) + U^0(P) + L[1 - PV(X^0)] \\ &= \frac{L[PV(X)] + FX + L[1 - PV(X^0)]}{1 - VX - U^0} \\ &= \frac{(1000)(.890) + 200 + (1000)[1 - .912]}{1 - .20 - .04} \\ &= 1,550 \end{aligned}$$

But note why I think there is an inconsistency in the two parts. We used the same pricing logic in both parts, but this premium implies a loss ratio of:

$$PLR = \frac{1000}{1550} = 64.5\%$$

This doesn't match the answer above. The reason is, I believe, that the 2% underwriting profit provision in the reviewed line's rate isn't consistent with the other information. For the two parts to be consistent the underwriting profit provision would have to be 2.59%.

In that case, the premium would be:

$$P = \frac{1000 + 200}{1 - .20 - .0259} = 1,550$$

Similarly, the Robbin formula used in the first question would tell us that:

$$PLR = \frac{.04 - .0259}{.9122 - .8904} = 64.5\%$$

This makes it hard to know what they were after in the first part. But hopefully this gives you a good sense for the PV Offset Method.

2009 Exam Question 14

An actuary has compiled the following information for a single line carrier:

Premium	10,000,000.00
Premium to Surplus Ratio	2.50
Policyholder supplied funds on unearned premiums	12.00%
3-year average ratio of loss reserves to losses incurred	1.20
Permissible loss ratio	63.00%
Tax Rate	35.00%
Traditional underwriting profit provision	10.50%
Final underwriting profit provision	3.20%

a. Using the Calendar Year Investment Income Offset procedure, calculate the after-tax portfolio yield for this insurer.

Recall that this method just uses the traditional underwriting provision and subtracts the investment income on policyholder supplied funds. The relationship between the traditional and adjusted underwriting profit provision is:

$$U = U_0 - i_{AT}(PHSF)$$

where,

$$\begin{aligned} PHSF &= \text{Unearned Premium Balance} + \text{Loss and LAE Reserves} \\ &= [\text{Unearned Premium Reserve Net of Prepaid Expenses} - \text{Premium Receivables}] \\ &\quad + [\text{Permissible Loss Ratio * (Ratio of Loss Reserves to Incurred Loss)}] \end{aligned}$$

We can calculate $PHSF$ from the information given and then plug into the formula for U to solve for the one unknown, which is the after-tax investment yield.

First, writing the $PHSF$ as a percent of premium we have:

$$PHSF = 12\% + 63\%(1.2) = 87.6\%$$

And then to solve for i_{AT} we have:

$$\begin{aligned} U &= U_0 - i_{AT}(PHSF) \\ 3.2\% &= 10.5\% - i_{AT}(87.6\%) \\ \Rightarrow i_{AT} &= (10.5\% - 3.2\%)/87.6\% = 8.33\% \end{aligned}$$

b. Calculate the after-tax investment income for this insurer.

The after-tax investment income is the after-tax investment rate, which we solved for above as $i_{AT} = 8.33\%$, multiplied by the total invested assets, which consist of the policyholder supplied funds and the surplus. We calculated the total $PHSF$ as being 87.6% of premium and we are told that the surplus is $1/2.5 = 40\%$ of premium, so the total after-tax investment income is:

$$\text{After-Tax Inv Income} = 8.33\%[(87.6\%)(10,000,000) + (40\%)(10,000,000)] = 1,063,333$$

c. Briefly discuss two reasons why the 3-year average ratio of loss reserves to loss incurred can sometimes be distorted

Changes in the mix of business, changes in the volume of business and changes in loss reserve adequacy can all impact this ratio.

2012 Exam Question 18

Using the present value cash flow return model, a new actuary determined the underwriting profit provision to be 5.5%. Given the following for a single policy:

- Loss = \$75
- Fixed Expenses = \$10
- Variable expense ratio = 24%
- Tax rate = 34%
- Premium to surplus ratio = 2.0
- Equity to surplus ratio = 1.3
- Investment yield = 10%
- Rate used for discounting = 8%
- Target return = 15%
- Losses are paid according to the following pattern at the end of each quarter: 35%, 25%, 25%, 15%.
- Premium and expenses are paid at policy inception
- Surplus equals investible equity and is assigned to the policy at inception and released at policy expiration
- Equity is assigned to the policy at inception and released at policy expiration

Demonstrate that the underwriting profit provision of 5.5% is consistent with the required underwriting profit calculated by the Present Value Cash Flow Return Model.

Under this method, the key is to calculate the present value (at an investment discount rate) of the underwriting cash flows, the investment income on the investible equity (the surplus in this case) and the taxes and setting this equal to the present value, at the target equity return, of the equity flows.

To simplify the tax calculations, we can just use the after-tax underwriting and investment cash flows.

Start by using the 5.5% UW profit provision to determine the premium:

$$P = \frac{75 + 10}{1 - .24 - .055} = 120.57$$

From this, we can get all of the UW cash flows, investment income and taxes as follows:

Time	Prem	Expense	Loss	UW CF	Inv Income	Total CF	After-Tax CF	PV(CF)
0.00	120.57	38.94	0.00	81.63		81.63	53.88	53.88
0.25			26.25	-26.25	1.51	-24.74	-16.33	-16.02
0.50			18.75	-18.75	1.51	-17.24	-11.38	-10.95
0.75			18.75	-18.75	1.51	-17.24	-11.38	-10.74
1.00			11.25	-11.25	1.51	-9.74	-6.43	-5.95
Total								10.21

In the table above, I used the 8% discount rate given in the question and just assumed that the investment income is earned at an annual rate of 10% on the committed surplus ($.5 * 120.57 = 60.285$) only.

From the above, we know the equity flows are 1.3 times the surplus flows and the present value of the equity flows is therefore:

$$PV(\Delta E) = 78.37 - 78.37/1.15 = 10.22$$

This shows that the PV of the cash flows is equal to the PV of the change in equity when the profit margin is 5.5%.

2014 Exam Question 17

Consider the following assumptions for a new policy:

Expected Claims	100,000
Expenses	20,000
Policyholder supplied funds as ratio to premium	43.60%
Tax Rate	35%
Before-tax return on invested assets	6%
Target return on equity	8%
Surplus to premium ratio	2.0
Surplus to equity ratio	1.0

a. Using the calendar year return on equity method, calculate the indicated price for the policy.

The Robin paper contains a formula for the underwriting profit provision (see below), but questions like this are easier without that formula. Just note that the method calculates an ROE using after-tax underwriting profit and after-tax investment income on all funds (policyholder supplied funds and capital) in the numerator. Writing the equation for the ROE in terms of the premium and setting this equal to the target ROE of 8% allows you to simply solve for the unknown premium:

$$\begin{aligned}
ROE &= \frac{\text{Underwriting Profit Provision in Dollars} + \text{Investment Income} - \text{Taxes}}{\text{Equity}} \\
&= \frac{U_{\$}(1-t) + \text{II}(1-t)}{EQ} \\
8\% &= \frac{(P - 100,000 - 20,000)(1 - .35) + [.436P + 2P](6\%)(1 - .35)}{2P} \\
\Rightarrow P &= 133,332
\end{aligned}$$

As mentioned above, the formula for the underwriting profit as a percentage of premium in the Robin paper could be used:

$$U = \frac{1}{1 - t_u}\left[(\text{Target ROE})\left(\frac{\text{E:S Ratio}}{\text{P:S Ratio}}\right) - i_{AT}\left(PHSF + \frac{1}{\text{P:S Ratio}}\right)\right]$$

From this, the premium would then be:

$$P = \frac{L + FE}{1 - VE - U}$$

b. Briefly describe two ways that risk is reflected in the calendar year return on equity method.

Risk is taken into account by both the selected surplus to premium ratio and the target ROE.

2010 Exam Question 9

An insurance company's most recent financial statement reflects the following:

Surplus	100
Premium	150
Reserves	200
Combined Ratio	105%
Investment Yield	6%

a. Calculate the total return on equity.

Using Ferrari's basic formula we have:

$$\begin{aligned}\frac{T}{S} &= \frac{I}{A}\left(1+\frac{R}{S}\right)+\frac{U}{P}\frac{P}{S}\\ &= 6\%\left(1+\frac{200}{100}\right)+(1-105\%)\frac{150}{100}\\ &= 10.5\%\end{aligned}$$

We could have also used Ferrari's alternative version:

$$\begin{aligned}\frac{T}{S} &= \frac{I}{A}+\frac{R}{S}\left(\frac{I}{A}+\frac{U}{R}\right)\\ \frac{T}{S} &= \frac{I}{A}+\frac{R}{S}\left(\frac{I}{A}+\frac{U}{P}\frac{P}{R}\right)\\ &= 6\%+\frac{200}{100}\left(6\%+(1-105\%)\frac{150}{200}\right)\\ &= 10.5\%\end{aligned}$$

b. Assuming the investment yield changes to 10%, calculate the surplus required to achieve a 16% total return on equity.

If we assume that changes in the investment yield don't have any effect on the other relationships given, this is just a matter of solving for S in the above formulas. The second version of the formula makes this algebra easier, so I will use that one.

$$\begin{aligned}\frac{T}{S} &= \frac{I}{A}+\frac{R}{S}\left(\frac{I}{A}+\frac{U}{R}\right)\\ 16\% &= 10\%+\frac{200}{S}\left(10\%+(1-105\%)\frac{150}{200}\right)\\ \Rightarrow S &= 208.33\end{aligned}$$

2009 Exam Question 8

The CEO of an insurance company seeks to increase the investment income through a higher investment allocation to alternative assets such as hedge funds. The company's investment manager indicates the new asset allocation would double the yield on investments, and the CFO believes that stockholder equity would need to be 50% above its current level.

The following table summarizes current financial data on the company:

Target Return on Stockholder Equity (T/S)	10%
Premium to Stockholder Equity (P/S)	2.5
Stockholder Equity to Assets (S/A)	0.25
Combined Ratio	104%

Stockholder Equity (S) is defined as the sum of the capital, surplus and the equity in the unearned premium reserves.

For the first approximation, the company decides to make the following assumptions:

- Stockholder equity is fully invested
- Taxes do not impact the calculation

a. The company implements the new investment strategy, raising the required additional funds needed with no change in premium or reserves. Calculate the combined ratio necessary to achieve the target return on equity.

From Ferrari's basic formula:

$$\frac{T}{S} = \frac{I}{A}\left(1 + \frac{R}{S}\right) + \frac{U}{P}\frac{P}{S}$$

Using this, and the information given in the question, we can calculate the unknown investment return on assets assuming that the current total return on equity is equal to the target return of 10%.

We were told that the combined ratio was 104%, so that means that the underwriting profit as a percent of premium is $U/P = 1 - 104\% = -4\%$.

The only other ratio we weren't given directly in the question was the reserve to surplus ratio. However, we were told that surplus is 25% of assets, which means reserves are 75% of assets and the reserve to surplus ratio is:

$$\frac{R}{S} = \frac{75\%}{25\%} = 3.0$$

Using this, we have the following that can be solved to find the investment yield:

$$10\% = \frac{I}{A}(1 + 3.0) + (-4\%)(2.5)$$

Solving this for the investment yield gives us $I/A = 5\%$.

Now we just need to adjust the ratios to reflect a doubling of the investment yield and a 50% increase in surplus.

- $I/A = 2 * 5\% = 10\%$
- $R/S = 3.0/1.5 = 2.0$
- $P/S = 2.5/1.5 = 1.67$

All we need to do now is calculate the ratio U/P to achieve a 10% return on equity.

$$10\% = (10\%)(1 + 2.0) + \frac{U}{P}(1.67)$$

This gives $U/P = -12\%$ and a combined ratio of 112%.

b. An analyst notes that an insurance operation is effectively borrowing funds from policyholders. Discuss the cost of this borrowing and whether it is fixed or variable.

To see the expected cost of this borrowing, rearrange the basic Ferrari equation as follows:

$$\frac{T}{S} = \frac{I}{A} + \frac{R}{S}\left(\frac{I}{A} + \frac{U}{R}\right)$$

The last term shows that the total return is increased by the leverage ratio (R/S), which reflects the amount of funds borrowed from policyholders in the form of reserves, multiplied by the net benefit of this borrowing.

The net benefit is the investment yield plus the underwriting profit as a percent of reserves. To the extent that the underwriting profit is negative, U/R can be thought of as a cost of the borrowed funds. If it is positive, then this borrowing has a negative cost.

However, unlike traditional borrowing at fixed costs, the underwriting profit is unknown upfront and actual underwriting results will determine what this costs ultimately is.

c. For the company described above, calculate the cost of borrowing (in percentage points) after the revised allocation.

The answer to this question depends on how you interpret the phrase “after the revised allocation”. The firs part of the question said that premiums and reserves are unchanged, which means that after the allocation the ratio of U/P is still -4% and our total return is now different from the 10% assumed initially. The fact that we solved for the ratio $U/P = -12\%$ needed to maintain the same total return should be irrelevant as we were not told that they made changes in line to do so.

So, we assume that after the changes the underwriting profit relative to premium is unchanged and so the formula above suggests that the total return becomes:

$$\begin{aligned}\frac{T}{S} &= \frac{I}{A}\left(1+\frac{R}{S}\right)+\frac{U}{P}\frac{P}{S}\\ &= 10\%(3)+(-4\%)(1.67)\\ &= 23.33\%\end{aligned}$$

To quickly get the underwriting profit relative to reserves we can plug this result into the alternative formula for the total return:

$$\begin{aligned}\frac{T}{S} &= \frac{I}{A}+\frac{R}{S}\left(\frac{I}{A}+\frac{U}{R}\right)\\ 23.33\% &= 10\%+(2)\left(10\%+\frac{U}{R}\right)\\ \rightarrow \frac{U}{R} &= -3.33\%\end{aligned}$$

2010 Exam Question 14

An insurance company currently writes policy X and is considering adding policy Y. The losses on the two accounts are represented below:

		Loss for Account	
Event	Probability	X	Y
1	1.0%	10,000.00	500.00
2	0.5%	30,000.00	20,000.00
3	4.0%	500.00	3,000.00

The variance risk load multiplier is 0.00004.

a. Calculate the marginal variance risk load for each of policy X and policy Y.

The first step is to calculate the variances for the portfolio containing just X and then for the portfolio containing X and Y. In this case we can use the binomial approximation for each event to get the following formulas:

$$\text{Mean} = \sum p(i)X_i$$

$$\text{Variance} = \sum p(i)[1-p(i)]X_i^2$$

$$\text{Covariance}(L,n) = \sum p(i)[1-p(i)]X_iY_i$$

The calculations for the variances are shown below:

	Mean			Variance		
Event	p_iX_i	p_iY_i	$p_i(X_i+Y_i)$	$p_i(1-p_i)X_i^2$	$p_i(1-p_i)Y_i^2$	$p_i(1-p_i)(X_i+Y_i)^2$
1	100	5	105	990,000	2,475	1,091,475.00
2	150	100	250	4,477,500	1,990,000	12,437,500.00
3	20	120	140	9,600	345,600	470,400.00
	270	**225**	**495**	**5,477,100**	**2,338,075**	**13,999,375**

Now to calculate the marginal variance risk loads, we use the variance of X as the marginal variance since it is the only risk in the portfolio at first. Then the marginal variance for Y is the difference between the total variance and the variance before adding it to the portfolio. The MV risk load is then the marginal variance multiplied by $\lambda = .00004$.

These calculations are shown below:

	X	Y	Total
Mean	270	225	495
Std Dev	2,340	1,529	3,742
Variance	5,477,100	2,338,075	13,999,375
Covariance	3,092,100		
Marginal Variance	5,477,100	8,522,275	
Lambda	0.00004	0.00004	
MV Risk Load	**219**	**341**	

b. Calculate the covariance share risk load for each of policy X and policy Y.

Mango showed that you can generalize the Shapley method to allow for the two risks to share their covariance in any number of ways. One in particular he recommended in this type of problem is to have them share the covariance in proportion to the expected losses *by event.*

To do this, we recalculate the covariance but apply an event-specific weight to the calculations as follows:

$$CovShare_X = \sum_{i=1}^{n} 2\frac{X_i}{X_i + Y_i} p_i(1 - p_i)X_iY_i$$

$$CovShare_Y = \sum_{i=1}^{n} 2\frac{Y_i}{X_i + Y_i} p_i(1 - p_i)X_iY_i$$

Event	Loss for X	Loss for Y	Weight X	Weight Y	Cov Share X	Cov Share Y
1	10,000	500	95.2%	4.8%	94,286	4,714
2	30,000	20,000	60.0%	40.0%	3,582,000	2,388,000
3	500	3,000	14.3%	85.7%	16,457	98,743
					3,692,743	**2,491,457**

But notice a subtle point. Although Mango was primarily concerned with pricing both risks on a renewal basis and therefore applying the covariance share amounts as the marginal risk measures in both cases, this question made it clear that X is an existing risk and so it isn't clear if they want the "build-up" or "renewal" risk load for X. I will show both below, but the answer I would have given is the build-up case, which is the same as in the previous part.

Finally, the risk loads are given as:

$$\text{Build-Up Risk Load}_X = \lambda \text{Var}(X) = \mathbf{219}$$

$$\text{Renewal Risk Load}_Y = \lambda[\text{Var}(Y) + CovShare_Y] = \mathbf{193}$$

$$\text{Renewal Risk Load}_X = \lambda[\text{Var}(X) + CovShare_X)] = 367$$

The last amount shown is just for information purposes. The answers in bold should be what they were looking for.

c. Calculate the deferred risk load.

The *deferred risk load* is the term Mango gave (only in the exhibit in the appendix, by the way) to the portion of the renewal risk load under the covariance share (or Shapley) methods that is **not** captured in the build-up risk load case. It represents the risk load attributable to the portion of the covariance between the risks that could not be included in the build-up risk load for the first risk because it wasn't known yet.

It is an important number because if we price the first risk without capturing this amount and then price fairly the other risks so that each one only has to be charged for its share of the covariance with the other risks we will wind up not charging enough in total. The first risk gets a break.

As we see in the results for the previous part, the difference between the build-up risk load for X and the renewal risk load for X under the covariance share method is $367 - 219 = 148$.

The numerical calculations of the difference between the build-up and renewal risk loads are as follows:

	X	Y
Build Up		
Stand-Alone Variance	5,477,100.00	2,338,075.00
CovShare		2,491,457.14
Change in Var	5,477,100.00	4,829,532.14
Lambda	0.00004	0.00004
Risk Load	219.08	193.18
Renewal		
Stand-Alone Variance	5,477,100.00	2,338,075.00
CovShare	3,692,742.86	2,491,457.14
Change in Var	9,169,842.86	4,829,532.14
Lambda	0.00004	0.00004
Risk Load	366.79	193.18
Difference	147.71	0

2009 Exam Question 12

A company's portfolio currently consists of a single account (X). The company is considering adding a second account (Y). The following variances of aggregate loss are known:

- Variance of X = $640,000$
- Variance of Y = $160,000$
- Variance of X+Y = $1,000,000$

The risk load multiplier is based on a required return on marginal surplus of 25% and a standard normal multiplier of 2.0.

a. Calculate the risk loads for account X, account Y and the total portfolio using the Marginal Surplus method.

To use this method we apply the following formula:

$$\text{MS Risk Load} = \frac{yz}{1+y}(S' - S)$$

where $(S' - S)$ represents the marginal *standard deviations* and the fraction represents the risk load multiplier:

$$\text{Risk Load Multiplier} = \frac{yz}{1+y} = \frac{.25(2)}{1.25} = .40$$

Now we calculate the marginal standard deviations. Risk X begins as the only risk, so it's marginal standard deviation is its stand-alone standard deviation. For Y, we calculate the marginal standard deviation as the difference between the total portfolio standard deviation and X's standard deviation. The risk load is then the .40 multiplier times these marginal amounts.

The calculations are shown below:

	X	Y	X+Y
Stand-Alone Std Dev	800	400	1,000
Marginal Std. Dev.	800	200	1,000
Risk Load Multiplier	40%	40%	40%
Risk Load	320	80	400

b. Calculate the risk loads for account X, account Y and the total portfolio using the Marginal Variance method.

The only differences here are that we use marginal variances rather than marginal standard deviations and, to be consistent, we scale the MS risk load multiplier by the total portfolio standard deviation:

$$\lambda = \frac{yz/(1+y)}{\text{Standard Deviation of } X+Y} = \frac{.40}{1000} = .00040$$

	X	Y	X+Y
Stand-Alone Variance	640,000	160,000	1,000,000
Marginal Variance	640,000	360,000	1,000,000
Risk Load Multiplier	0.00040	0.00040	0.00040
Risk Load	256	144	400

c. Upon renewal, explain what would happen to the total portfolio risk loads calculated in each of parts (a) and (b) above.

Upon renewal, the calculation of the marginal surplus and the marginal variance for Risk Y would stay the same but the calculation of both of these amounts for Risk X would change because it would now reflect the existence of Risk Y in the portfolio.

In the case of the MS method, the marginal standard deviation for X will be lower than the stand-alone standard deviation and so the risk load will decrease. For the MV method, the marginal variance will be larger than the stand-alone variance (because it will include two times the covariance of X and Y, which wasn't captured before) and so the risk load will increase.

The following tables show the calculations, which don't appear to be needed for the exam answer:

Marginal Surplus Method			
	Renew X	Renew Y	Total
Stand-Alone Std Dev	800	400	
Marginal Std. Dev.	600	200	
Risk Load Multiplier	40%	40%	
Risk Load	**240**	**80**	**320**
Build-Up Risk Load	320	80	400
Difference	-80	0	-80

Marginal Variance Method			
	Renew X	Renew Y	Total
Stand-Alone Variance	640,000	160,000	
Marginal Variance	840,000	360,000	
Risk Load Multiplier	0.00040	0.00040	
Risk Load	**336**	**144**	**480**
Build-Up Risk Load	256	144	400
Difference	80	0	80

2008 Exam Question 11

A property catastrophe insurer is currently insuring two accounts, both nearing their renewal date. The following information is available:

Event	Probability	Loss for Account X	Loss for Account Y	Total Loss	Covariance Share X	Covariance Share Y
1	0.020	32,000.00	7,500.00	39,500	7,621,670.89	1,786,329.11
2	0.020	13,000.00	5,000.00	18,000	1,840,222.22	707,777.78
3	0.030	5,000.00	11,000.00	16,000	1,000,312.50	2,200,687.50
4	0.015	14,000.00	6,000.00	20,000	1,737,540.00	744,660.00
5	0.005	1,000.00	29,000.00	30,000	9,618.33	278,931.67
6	0.010	1,000.00	2,000.00	3,000	13,200.00	26,400.00
7	0.010	10,000.00	5,000.00	15,000	660,000.00	330,000.00
Total					12,882,564	6,074,786
	Var(X)	28,011,075.00	Var(X+Y)	57,085,000.00		
	Var(Y)	10,116,575.00	Cov(X, Y)	9,478,675.00		
	Risk Load Multiplier	0.000039				

a. Using the Shapley Method, calculate the risk load for both Account X and Account Y.

On a renewal basis, each account is priced by treating the marginal risk as being the sum of the stand-alone variance and the covariance. By doing this, we force each account to share their covariance equally and then we are assured, even in a renewal sense, that the total portfolio risk load is achieved, a property known as *renewal additivity*.

$$\text{Shapley Value for Marginal Risk from Renewing } X = \text{Var}(X) + \text{Cov}(X, Y)$$

$$\text{Shapley Value for Marginal Risk from Renewing } Y = \text{Var}(Y) + \text{Cov}(X, Y)$$

Plugging in the values provided:

Shapley	Marginal Risk	Risk Load
X	37,489,750	1,462
Y	19,595,250	764

b. Using the Covariance Share Method, calculate the risk load for both Account X and Account Y.

On a renewal basis, each account is priced by treating the marginal risk as being the sum of the stand-alone variance and its respective covariance share, which is essentially a weighted allocation of the contribution of the covariance to the total risk. By doing this, we force each account to share their covariance in proportion to their respective losses by event. This too is renewal additive, but perhaps a more fair allocation between risks.

Covariance Share	Marginal Risk	Risk Load
X	40,893,639	1,594
Y	16,191,361	631

2012 Exam Question 13

An insurance company writes homeowners accounts with the following exposures to independent catastrophe events: They have a target return on equity of 15%.

Event	Probability	Account A Loss	Account B Loss	Total (A+B)
1	0.1%	200	100	300
2	1.0%	100	50	150
3	2.0%	80	60	140

a. Calculate the renewal risk load for each account using the marginal surplus method, targeting a 2% probability of ruin.

Begin by noting that for the marginal surplus method, the renewal risk loads are determined by multiplying the marginal standard deviation as if the account were the last one in the portfolio by a risk load multiplier equal to:

$$\text{Risk Load Multiplier} = \frac{yz}{1+y}$$

where $y = 15\%$ in this problem and $z = 2.054$ based on the z-score for a standard normal distribution at the 98th percentile.

To get the marginal standard deviations, we use the event loss table given to find the standalone and combined standard deviations, as follows:

Standard Deviation for Account A				
Event	Probability	1-Prob	Account A Loss	$p(1-p)L^2$
1	0.1%	99.9%	200	39.96
2	1.0%	99.0%	100	99.00
3	2.0%	98.0%	80	125.44
			Variance	**264.40**
			Std Dev	**16.26**

Standard Deviation for Account B				
Event	Probability	1-Prob	Account B Loss	$p(1-p)L^2$
1	0.1%	99.9%	100	9.99
2	1.0%	99.0%	50	24.75
3	2.0%	98.0%	60	70.56
			Variance	**105.30**
			Std Dev	**10.26**

Standard Deviation for Portfolio of Account A and B				
Event	Probability	1-Prob	Total (A+B)	$p(1-p)L^2$
1	0.1%	99.9%	300	89.91
2	1.0%	99.0%	150	222.75
3	2.0%	98.0%	140	384.16
			Variance	**696.82**
			Std Dev	**26.40**

Finally, the tables below show the calculations of the renewal risk loads:

Risk Load for A		Risk Load for B	
Std Dev Total	26.40	Std Dev Total	26.40
Std Dev B	10.26	Std Dev A	16.26
Marg Std Dev A	16.14	Marg Std Dev B	10.14
Risk Load Multiplier	0.27	Risk Load Multiplier	0.27
Risk Load	4.32	Risk Load	2.72

b. Explain whether a company should or should not expect to achieve its target return on equity using the risk loads from part (a) above.

Under the marginal surplus method, renewal risk loads are sub-additive and therefore the total risk load will not add up to the marginal surplus risk load required for the whole portfolio.

c. Explain how the answer to part (b) above would change if the marginal variance method of calculating the risk loads were used.

In this case, the renewal risk loads will add up to more than the risk load required for the whole portfolio as the result of double-counting the covariance in the marginal variance calculations.

d. The company uses the CAPM framework to set target return on equity. Explain the impact on each accounts renewal risk load under the marginal surplus method if the market risk premium increases.

This will increase the required return, y, and therefore increase the risk load multiplier and the risk load.

2009 Exam Question 13

A company is considering adding a new insured to its portfolio and has compiled the following information:

- Current surplus = \$100 million
- Std deviation of losses for current portfolio = \$ 40 million
- Expected return on current portfolio = \$20 million
- Standard deviation of losses for proposed portfolio = \$45 million
- Desired return on marginal surplus for new policies = 15%

a. Calculate the minimum expected return in dollars for the new policy to produce the desired return on marginal surplus (using the Kreps marginal surplus risk load method presented in the Mango reading).

Recall from the notes (see the Risk Loads Background reading on page 683) that under Kreps' marginal surplus method the capital (surplus) contributed, V, is equal to some multiple of the standard deviation less the expected profit (in dollars), or

$$V = zS - R$$

where z is a multiple of the standard deviation which results in some target (small) probability of ruin, S is the standard deviation of the existing book of business and R is the expected profit of the existing book of business.

Writing an additional policy, with expected profit (i.e. risk load) r, standard deviation σ and correlation with the existing book C, we can write the *new* surplus, overall standard deviation and expected profit as V', S' and $R' = R + r$. This gives the following for the marginal surplus required:

$$V' - V = z(S' - S) - r$$

Then, the risk load is simply a target return (over the risk free rate) on this surplus so that investing surplus to write this policy is at least as attractive as investing the surplus in other marketable securities with similar risk.

Denoting this target return as y we have:

$$\begin{aligned} r &= y(V' - V) \\ &= y[z(S' - S) - r] \\ &= \frac{yz}{1+y}(S' - S) \end{aligned}$$

Now, in this question we were really just asked to apply this method and were given the value for $y = 15\%$, the value of the marginal standard deviation, $S' - S = 45 - 40 = 5$ million. We were asked to solve for the dollar profit, r for the new policy but were not given the value for z.

But we can solve for the value of z that is consistent with the existing portfolio. We were told the existing portfolio's standard deviation, S; it's expected profit, R, and the current surplus, V. Therefore, we can solve for z in the following:

$$\begin{aligned} V &= zS - R \\ 100 &= z(40) - 20 \\ &\Rightarrow z = 3.0 \end{aligned}$$

Now, solving the main question is trivial:

$$\begin{aligned} r &= y[z(S' - S) - r] \\ &= .15[3(5 \text{ million}) - r] \\ &= .15[3(5 \text{ million}) - .15r] \\ 1.15r &= 2.25 \\ r &= 1.96 \text{ million} \end{aligned}$$

2010 Exam Question 8

An insurer writes one policy that makes a single loss payment one year after the collection of premium.

Taxes and Expenses	0
Risk-free Rate	3%
Target Return on Equity	12%
Expected Loss	80
Loss Safety Level	200

Assuming that assets are invested in a risk-free bond, use the swap method to calculate the risk load.

In the swap case, Kreps' formula for the risk load was simply:

$$R = \frac{y - r_f}{1 + r_f} A$$

where y is the expected return on a risky investment, r_f is the risk-free rate and A is the amount of surplus invested by the shareholders. The question didn't actually provide the value for y, it provided the target return on equity or IRR. But in the Kreps method, a key step is to set the IRR and the risky investment yield to be the same, so the test writers appear to have meant for you to assume $y = 12\%$.

They also provided the value of r_f but we need to determine the *maximum* that A would be under two sets of constraints (so that both constraints are satisfied).

The first constraint is that assets (the sum of premiums and A) are sufficient to fund losses to the loss safety level, s. This gives us the following constraint:

$$\begin{aligned} A &\geq \frac{s - \mu_L}{1 + y} \\ &\geq \frac{200 - 80}{1.12} \\ &\geq 107.14 \end{aligned}$$

The second constraint is that the variance of the reinsurance and swap strategy together is less than the variance of having invested all of A directly in the risky investment. Unfortunately, the question didn't provide any information about the variance of the losses or the risky investment, so we can only assume that the safety constraint will be the binding one and take $A = 107.14$.

Now that we have all of the inputs, we can plug the values into the formula we had before for the risk load:

$$\begin{aligned} R &= \frac{y - r_f}{1 + r_f} A \\ &= \frac{.12 - .03}{1.03}(107.14) \\ &= 9.36 \end{aligned}$$

2009 Exam Question 10

The following information is available for a potential insured:

- The insured is assumed to have a loss distribution as follows:

$$F[L] = \begin{cases} 0 & 0 \le L < .8 \\ U[0, 1000] & .8 \le L \le 1 \end{cases}$$

  where $U[a, b]$ represents the uniform distribution from a to b.

- The standard deviation of this distribution is \$231.

- The distribution of the target investment return has a mean of 15% and a standard deviation of 30%.

- The total assets available from investment need to fund one 50-year loss event.

- The risk-free rate is 5%.

- All losses are paid one year after policy inception. *Note that this wasn't in the original question but is important because it affects the results.*

a. Use Kreps' Investment Equivalent Reinsurance Pricing methodology to calculate the premium, risk load and assets needed to support the insured using solely his Swap financing strategy.

Note that I have slightly altered the wording of this question to narrow the scope.

Recall that Kreps recommended using two different financing options to calculate the risk load and to select the *smaller* of the two results. Here, there was really nothing in the question that prevented you from applying both strategies other than the excessive amount of calculations required to apply the Put Option strategy. This is why I added the language to use only his Swap strategy.

In the Swap case,

$$R = \frac{y - r_f}{1 + r_f} A$$

To calculate the amount of assets, A, we consider two constraints — the safety constraint and the variance constraint.

Under the safety constraint,

$$A \ge \frac{s - \mu_L}{1 + y}$$

where s is the dollar amount of loss at the safety level, which is the one-in-50 year loss level or the 98th percentile ($1 - 1/50 = 98\%$). In this case, despite the odd way in which the loss distribution was presented, losses are uniformly distributed between \$0 and \$1000 for cumulative probabilities between 80% and 100%. This means that we can find s as:

$$s = \frac{.98 - .80}{1.00 - .80}(1,000) = 900$$

And similarly, we can calculate the mean as $\mu_L = .8 * (0) + .2 * (500) = 100$. This results in the value of assets at the equality of:

$$A = \frac{900 - 100}{1.15} = 696$$

Under the variance constraint we use the formula:

$$A \geq \frac{\sigma_L}{\sigma_y} = \frac{231}{30\%} = 770$$

Since the variance constraint is binding, we use $A = 770$ and calculate the risk load as:

$$R = \frac{y - r_f}{1 + r_f}A = \frac{.15 - .05}{1.05}(770) = 73$$

This gives a final premium, assuming no expenses, of:

$$P = PV(L) + R = \frac{100}{1.05} + 73 = 168.57$$

b. Explain whether the loss safety constraint or the investment variance constraint is dominant under these conditions.

It was shown above that the variance constraint required more assets, so this is the dominant constraint.

c. The actuary is told that the total assets available for investment now need to fund a 250-year loss event, rather than a 50-year loss event. Explain whether this change of constraint will alter the answer to Part (b) above.

We would normally expect that increasing the loss safety level significantly into the tail of the distribution will make the safety constraint dominant because this change does not impact the variance constraint at all. That is entirely dependent on the loss and investment return standard deviations. But in this case, because the losses were uniform rather than skewed, the effect of a higher return time (250-year loss vs. 50-year loss) is muted.

We can check this directly by recalculating the assets required under the safety constraint as:

$$s = \frac{(1 - 1/250) - .80}{1.00 - .80}(1,000) = 980$$

This leads to $A = 765$. As noted before, the variance constraint is unaffected and o the assets required are still 770 and the variance constraint is still dominant.

2012 Exam Question 15

An actuary has been asked to determine the risk load for a reinsurance contract that will make a single loss payment at the end of one year. If the contract is written, the reinsurer will invest the premium and any other allocated assets.

Given the following information:

- Risk free rate = 2%
- Yield of current investment portfolio = 5%
- Target return on equity = 12%
- Funds invested by reinsurer = $4,000,000
- Mean value of loss for the contract = $3,000,000
- Standard deviation of loss for the contract = $540,000

a. Calculate the reinsurance premium using Kreps' swap method.

In the Kreps swap strategy we plan to collect premiums, P, and invest our own assets, A, such that at the end of the year our risk free investment after paying the claim amounts, L, will grow to $(1 + r_f)(P + A) - L$. Since our initial investment is A, the IRR of this combined reinsurance and financial strategy is given as:

$$(1 + \text{IRR})A = (1 + r_f)(P + A) - L$$

where the total amount invested in the risk free assets is denoted as $F = P + A$.

In the Kreps paper, he specifically set the IRR equal to the yield on a risky asset portfolio, which presumably in this question would be 5%. But here they gave you a target return of 15%. The Examiner's Report suggests that they expected you to use this target return for the IRR rather than the investment yield, which is technically not what Kreps did. Given slightly different wording, this would have been a reasonable question to ask, but they should not have referenced the Kreps paper.

Using this target return on the invested assets, we can rewrite the previous formula as:

$$(1 + y)A = (1 + r_f)(P + A) - L$$

To get the formula for the risk load, R, we substitute in for the premium so that P is equal to the risk free present value of the expected claims plus the risk load R:

$$(1 + y)A = (1 + r_f)\left[\frac{L}{1 + r_f} + R + A\right] - L$$

and then we can solve for R:

$$R = \frac{y - r_f}{1 + r_f} A$$

In the Kreps method he established two constraints, the safety constraint and the variance constraint. These are used to help determine the amount of initial assets that need to be invested, A. In this problem, the total funds invested (the sum of the premium and the initial assets) have already been given as equal to 4,000,000, so we can just plug in that value and solve for A quite easily:

$$\begin{aligned}(1 + y)A &= (1 + r_f)(4,000,000) - 3,000,000 \\ A &= \frac{(1 + r_f)(4,000,000) - 3,000,000}{1 + y} \\ &= 964,286\end{aligned}$$

Plugging that into the formula for the risk load, we have:

$$\begin{aligned}R &= \frac{y - r_f}{1 + r_f} A \\ &= \frac{.12 - .02}{1.02}(964,286) \\ &= 94,538\end{aligned}$$

Adding in the discounted expected loss amounts gives a total premium of:

$$P = 3,000,000/1.02 + 94,538 = 3,035,714$$

b. Describe what the loss safety level represents and what implications a higher loss safety level has for a reinsurer.

The loss safety level is a constraint on the probability of not having sufficient funds to pay the claims. It is used to establish a minimum level of shareholder assets that are needed to support writing the policy. If the safety level is higher, more funds are needed and therefore a higher risk load is needed so that the target rate of return can be earned on the larger fund balance.

c. Suppose the assumption of a single loss payment is relaxed, allowing multiple loss payments. Design a simple process to determine a risk load using stochastic modelling techniques.

This part of the question was from a part of the reading excluded from the syllabus. Even though this question wasn't graded, a quick version of the answer might still be useful.

When there are multiple payments, Kreps recommended simulating the loss amounts, investment income and asset balances to fully take into account their interactions. Using an initial guess at a risk load the simulation results can be used to confirm that the safety constraint and the variance constraint are both satisfied and that the average return equals the target return. If not, modify the risk load and repeat the simulation. Continue until the risk load that meets the return target and satisfies both constraints is found.

Appendix: Normal CDF Table

TABLE 1. Cumulative Normal Distribution (Positive x)

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 2. Cumulative Normal Distribution (Negative x)

x	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000