Jun Pan
junpan@mit.edu

253-3083
E62-624

## Class 1: Introduction

This Version: September 8, 2016

I'll write a note for each class. As you will see, these lecture notes are often more conversational, not as formal as a textbook. By allowing myself to be flexible and spontaneous in writing these notes, I actually get a lot of joy by just writing. Hopefully, they bring some joy to your reading as well. As of July 2016, I have been at MIT Sloan for 16 years. Teaching is not something that came naturally to me and I had my ups and downs. But one thing I know very well by now is that I enjoy teaching the most when I am sharing — knowledge, excitement, amazement, rational analysis, calm calculation, and quiet appreciation. So if I was not able to convey these fully in a classroom setting, I hope to use these lecture notes to get a second chance.

# 1 What to Expect from This Class?

- **This is not going to be a dry or dull class** because, as a subject matter, Finance is just too exciting to be taught in a boring way. Finance and financial markets are in real time and their ups and downs affect most people's lives.

  Last August, I was in Trout Lake, WA for a retreat. I thought I was as far away from Wall Street as possible. The location was rural America at its best and my fellow retreat participants are your typical spiritual type whose resume will not pass the first round of screening for any Wall Street jobs. Not that they are interested in such a job anyways. And yet, when the teacher mentioned in passing that the Dow had just lost 1,000 points that morning, there was a wave of agitation in that secluded woods we were in. "Everything was just fine before he had to mention that!" The woman sitting in front of me turned around and shared her thoughts during the break. I thought she was joking and was about to add mine when she murmured to herself, "Now I have to worry about my retirement money."

  In fact, this was not the first time when exciting financial news was broadcast to me in a retreat. When the markets went crazy in August 2011 because of the fear of

contagion of the European debt crisis, I was in a meditation retreat in Serpentine, Australia. It was as far away from any known financial centers as you can imagine and we were not allowed to have Internet access. One day, I was mindfully doing my walking meditation, and the caretaker of the retreat center literally stopped me in my tracks and broadcast the news to me. It was only much later did I learn that he was doing currency speculation on the side and had just lost a lot of borrowed money that day.

For the famous week of Lehman bankruptcy (Monday) and AIG bailout (Tuesday), I was again in a meditation retreat. This time, I was not broadcast any news – this retreat center is known for its strictness. One week later, I was in Boston, digesting the shock. My conclusion: I should not be going to any retreat anymore. But more seriously, these examples bring home to us the widespread impact of the financial markets. It would be a pity to view Finance simply as formulas made up of Greek letters.

- **I'll teach this class as a professor, not as a professional investor.** Finance as an academic discipline has played and continues to play an important role in how Finance is practiced in real life. It builds theoretical models and frameworks, through which the seemingly random events in the financial markets can be analyzed, understood and quantified. It offers pricing models to facilitate the tremendous innovations in financial products. More often than not, the most creative ideas and the best trading strategies arise from research papers written by Finance professors. In the early days of the 70s, such Finance professors were called "academia nuts." Today, Finance professors are often heavily sought after and their research papers carefully studied and followed by practitioners. MIT Sloan is highly respected in Wall Street, not because of the professional investors we helped to produce, but because of the academic work done by Paul Samuelson, Franco Modigliani, Bob Merton, Fischer Black, Myron Scholes, and many other professors.

  Over the years, I enjoy many conversations with professional investors. Their anecdotes are often fun and exciting and one can easily spend an hour talking about a particular trade or event. Such anecdotes add more texture and color and will be used periodically in the class. But without a systematic framework, anecdotes remain just anecdotes. I know that many students are eager to get into the "real world" and shun the classroom materials as being too academic. Well, let me be honest here. Sooner or later, you are going to be in the "real world." Where else? And you will spend your days being buried in anecdotes and events. What is the hurry? Being in school is a rare opportunity in

your life to retreat from the world and observe the events from a higher vantage point. Instead of being so eager to jump into the sea of uncertainty, why not first learn some basics about navigation?

- **This class will be an empirically driven class.** Over the years, Finance has become an empirically relevant discipline. Just imagine the amount of financial data that a Bloomberg terminal has to endure second by second, or millisecond by millisecond. When I was a PhD student in the late 1990s, I was given a quota of several gigabytes on a Sun Sparc workstation and I was thrilled. Today, that several gigabytes can hold maybe several minutes of stock trading data. How do we make sense of this jungle of data? Indeed, with the increasing availability of data and computing power, researchers have started to test the models developed in the 60s and 70s. By late 1990s, empirical work has replaced theoretical work as a more active research field in Finance. Studying the financial markets through the combination of data and theory is something I feel passionate about in my research. I find it exciting to be able to extract information from the seemingly noisy financial data. Analyzing and quantifying the regularity of financial risks from uncertain events is something I enjoy doing as a researcher. And I hope to be able to pass on this passion to you and help you develop these empirical skills, in addition to teaching you the basics of the financial markets. As such, this class will be an empirically driven class. We will talk theory, but theory is more of a guideline, like a map. This class is not about studying the map. It's about walking the path.

- **All class materials are available on Stellar.** There is not a required text book for this class. You can use Bodie, Kane, and Marcus as a reference. I will post the slides prior to each class. If you would like to take notes in the class, it might be a good idea to print a copy and bring it to the classroom. I will do my best to write a companion note such as this one for each class.

# 2 What do I Expect from the Students?

- **Come to the classroom with a love for the subject matter.** When I first realized that I could in fact make a career out of Finance, I was so exited that I put up a giant poster of the NYSE's trading floor in my tiny apartment. At the time, I was living in Brooklyn Heights and studying at NYU for my PhD in Physics. On my daily walk home via the Brooklyn Bridge, I often turned around at the end of the bridge to look

at the lights in lower Manhattan, imagining myself to be in one of the office buildings. Looking back, it was all so laughable. The NYSE trading floor is definitely not where the action is taking place and I had never even been close to a Wall Street career. But that love for the subject matter is what really matters. Over the years, I have read many of the Finance books written for the popular press, like *Capital Ideas* by Peter Bernstein, *When Genius Failed* by Roger Lowenstein, *Liar's Poker* and *Flash Boys* by Michael Lewis, Fool's Gold by Gillian Tett. Many of the books, I simply cannot put them down after reading the first few pages. By contract, I have never read one popular Physics book even though I also have a PhD in Physics. I read Physics books only to prepare for exams. I often joke with my students that your real passion is reflected by the books on your night stand.

So if you read Finance books only for exams, maybe this is not your cup of tea. So why not quit Finance and look for your real passion? You are still young and have most of your life ahead of you. Why get yourself stuck in a career that you do not love? For those of you who have to take this class in the meanwhile, I will suggest that you fake it until you make it, or until the end of the semester, whichever comes first.

- **Be mentally present in the classroom.** Each morning when you get up, you might wash yourself, put on some clean clothes, and make yourself presentable. Have you ever thought about how to prepare your mind? When you eat, you might watch your diet and be careful with what you put into your mouth. Have you ever thought about what you feed to your mind? Your mind is the best instrument in your life, and yet, without proper training, your mind is also the most vulnerable, incapable of fending off the waves of distractions in this digital age. If you ever wonder why you are not doing well in the class, in an interview, or in your job, it is possible that you have not prepared your mind well. Being smart with a high IQ comes mostly from one's genetic inheritance. It is over-rated anyways. Instead, be impressed by people in your life who are mentally present and aware. They carry themselves differently and they are usually light-hearted, flexible, and happy.

  Being mindful is something you can develop. Why not start your practice in the classroom? Think of the classroom as a sanctuary, away from the digital distractions that permeate our society. Do you know that some people pay a lot of money to attend digital detox retreats just to get away from the Internet and Cell connections for a week? Here, you can do it for free. Just refrain from the Internet for 80 minutes. Trust me, the rest of the world will be just fine without your digital footprint. They probably won't even notice. Give yourself a break and give me your full attention

during those 80 minutes. It will be a very good training of your mind.

- **As a general rule, all electronic devices including laptops and iPad are out.** If you really really need to use them, please talk to me. Please turn off your cell phones, including the *ding dong* sound for messages. If your phone is on vibration, please avoid putting it on hard surfaces. Otherwise, it will be just as noticeable.

- **Assignments and exams.** There are four group assignments, to be done in groups with no more than four students. Each assignment must be handed in before 5pm of the due day. Late assignments will not be accepted. The midterm exam will be given on Tuesday, October 18. The final exam will be given during the final exam week. There will be optional recitations held by TAs for the assignments and exams.

- **Let's keep our classroom a friendly environment.** This semester, we are going to spend 80 minutes together each time for 24 times. I would like us to create a friendly environment for one another. If one student happens to come into the classroom late, please try not to give him that "how dare you" look. Maybe he has a very good reason for being late. Who knows? Of course, this does not give you a free license to be late. You will certainly hear from me if I feel that you're consistently late. In this classroom, I would like everyone to feel comfortable enough to speak up, either for clarification or discussion. The only thing I would discourage in the classroom is students talking amongst themselves.

# 3 Modern Finance

After the 2008 financial crisis, I had the following conversation with my sister who works for a Pharmaceutical company. "Other than making money for themselves, what do people on Wall Street really do for the society?" She asked and I was quiet. "You know, everyday I go to work, I know that I am helping develop a drug that might help relieve people's pain, and I feel good about it. What about people on Wall Street?" She continued and I remained quiet. But inside, I had this huge question mark hanging over my head. To be honest, I have never fully resolved this doubt for myself. In 2008, the stock market dropped by 37%, and my passion for Finance was cut by just as much, if not more. What disappointed me was not the financial performance but the human performance in the financial industry.

In writing up the following account on the development of Modern Finance, I hope to be able to respond, at least partially, to my sister's question. Assuming that you will be part of the Wall Street in a few years, this should be a relevant question for you as well.

- **Markowitz (1952)** The beginning of Modern Finance can be dated by Markowitz (1952). As described in vivid details by Peter Bernstein in *"Capital Ideas,"* Harry Markowitz was a 25-year-old graduate student at Chicago, working on his PhD thesis. In 1990, he was awarded a Nobel Prize for his "pioneering work in the theory of financial economics."

  From the vantage point of today's knowledge base, the paper's insight is obvious and well understood by most people. First, Markowitz made the observation that, for any mean-variance investors, there is a risk and return tradeoff. Then, as any good researcher would do, he asked, given this tradeoff, what is this investor's optimal allocation to risk? "The answers Markowitz developed to these questions ultimately transformed the practice of investment management beyond recognition. They put some sense and some system into the haphazard manner in which most investors were assembling portfolios. Moreover, they formed the foundation for all subsequent theories on how financial markets work, how risk can be quantified, and even how corporations should finance themselves."[1]

  Against the backdrop of a single-minded focus on return at his time, Markowitz made the key insight that risk is central to the whole process of investing. In this day and age, any statement contrary to that observation is laughable. Risk is the single most important factor in Finance. No risk, no Finance. Financial markets, along with the tremendous innovations since 1970s, are vehicles designed to help us deal with risk.

  Well, this story does tell us how far Finance has evolved over the past half-century, with this one single insight made in the arcane world of academics by someone who had no direct interest or involvement in the stock market.

- **Tobin (1958)** Markowitz's insight was not recognized right away. After its initial publication in *the Journal of Finance*, the paper remained in obscurity for nearly ten years, attracting fewer than twenty citations in the academic literature. One of these citations was by James Tobin, a 1981 Nobel Prize winner. Tobin (1958) gave us the elegant result of two-fund separation. For any mean-variance investors, the optimal allocation consists of only two funds: one risky and one riskless. Regardless of their varying levels of risk aversion, all mean-variance investors hold exactly the same risky portfolio. The more risk-adverse investor allocates a smaller percentage of his wealth to the risky portfolio, but the composition of the his risky portfolio is exactly the same as everyone else.

[1] Chapter 2 of *"Capital Ideas,"* by Peter Bernstein.

This result gives us the striking insight that instead of getting lost in the sea of individual stocks, one should pay attention to this optimal risky portfolio. In today's world, with the increasing popularity of Index funds and ETFs, this idea seems quite obvious. But it was not until 1971, when the first Index Fund was created by John McQuown and his colleagues at Wells Fargo. And it was not until 1975, when the first Index Mutual Fund was created by John Bogle. This fund, now called the Vanguard 500 Index Fund, started off with just $11.3 million, a 93% shortfall from the initial target of $150 million. By 2014, over $2 trillion is invested in index mutual funds, accounting for 20% of the total net assets of equity mutual funds. From 2007 through 2014, index domestic equity mutual funds and ETFs received $1 trillion in net flows while actively managed domestic equity mutual funds experienced a net outflow of $659 billion.[2]

Again, the influence from the academic world is unmistakable. If you read the stories surrounding the creations of these index funds, you will see that these pioneers in industry were often influenced at a personal level by a few professors. Their convictions were often strengthened by the intellectual power behind the academic research. John Bogle writes, "Nobel laureate economist Paul Samuelson played a major role in precipitating the index fund's creation... Samuelson was much more forceful, strengthening my backbone for the hard task that lay ahead: taking on the industry establishment. His article 'Challenge to Judgment' caught me at the perfect moment. Published in the inaugural edition of the Journal of Portfolio Management in the autumn of 1974, it pleaded that some large foundation set up an in-house portfolio that tracks the S&P 500 Index ... ."[3]

- **Sharpe (1964)** If you ask me to pick one model in Finance that has the biggest and the longest-lasting impact, it will be the CAPM. Following the stream of Markowitz (1952), Tobin (1958), and Sharpe (1964), one has the reaction that this sequence of intellectual development is so natural that it is inevitable. But the last step in this "Investment Trilogy" is truly a giant leap. In 1990, Bill Sharpe was awarded a Nobel Prize for his "pioneering work in the theory of financial economics."

  In Markowitz (1952), the attention is on an individual investor. How much he should include a stock in his portfolio is determined by this stock's contribution to the risk (variance) and return (mean) of his portfolio. As such, what matters are the correlations between this stock and all other existing stocks in the portfolio. In the CAPM, the attention is on the entire economy. If every investor behaves optimally according

[2]The 2015 Fact Book by Investment Company Institute (ICI). Posted on Stellar under Readings.
[3]"How the Index Fund Was Born," by John C. Bogle, *Wall Street Journal,* 2011.

to the calculation in Markowitz (1952), what happens to the entire market when you aggregate this optimal individual behavior? Collectively, the markets should also clear: borrowing and lending in the riskfree market must net out and the entire wealth of the economy must be 100% allocated to the risky portfolio. In the academic language, the CAPM is a result of taking the partial-equilibrium model of Markowitz (1952) to equilibrium, a beautiful insight from Economics.

In equilibrium, the optimal risky portfolio in Tobin (1958) becomes the market portfolio – the single most important factor in the economy. In such an economy, you no longer have to keep track of the correlations of one stock with respect to all other stocks. What matters is a stock's correlation with the market portfolio. Hence $\beta_i = \text{cov}(R_i, R_m)/\text{var}(R_m)$. In this way, CAPM further clarifies the concept of risk. In particular, risk is not measured by the variance of an individual stock. For two stocks with the same variance, the one that comoves a lot with the market portfolio is the more risky one. Why? Because risk that is not correlated with the market portfolio can be diversified, but there is no way to diversify away the risk in the market portfolio. This concept of systematic risk is by far the most important intellectual development in Finance.

After singling out the systematic risk, the equilibrium analysis gives us the elegant pricing result. Simply put, you get paid for bearing the risk that matters:

$$E(R_i) - r_f = \beta_i \left(E(R_m) - r_f\right) .$$

If a stock contains purely idiosyncratic risk with $\beta_i = 0$, then you do not get paid for bearing this risk and the expected return is the same as the riskfree rate: $E(R_i) = r_f$.

- **Black and Scholes (1973)** In the 1970s, Fischer Black, Myron Scholes and Bob Merton did their pioneering work on Continuous-Time Finance at MIT Sloan, on the second floor of E52, I was told. In 1997, Merton and Scholes were awarded a Nobel Prize for "for a new method to determine the value of derivatives."

  The impact of this work is such that taking it out would be like switching off a bright light and the world of Finance would be a dim field due to its absence. Many of the financial markets we are going to study in this course would not have been created without this work. Even if such markets were in existence, people in these markets would be having a hard time figuring out how to price the product or hedge the risk.

  Going back to my sister's question, "What does Finance do for the society?" Finance

helps people deal with risk. No risk, no Finance. Markowitz (1952) gives us a framework to quantify the risk and return tradeoff. Sharpe (1964) points out that not all risk is equal and only systematic risk should be compensated. The work by Black, Merton, and Scholes takes this business of risk to a whole new dimension. When you purchase a stock, you get the entire package of risk that is inherent in this stock. In the language of academic Finance, you have a linear position and you own the entire distribution of the risk, both up and down. What if you don't want the entire distribution? What if you are interested in taking only some of the risk but not all? The financial innovation inspired by the work of Black, Merton, and Scholes is all about giving you more flexibility in dealing with risk.

Every summer, I go back to Shanghai to spend a month and half with my parents. In May 2015, before I was able to purchase the air ticket, I had to wait for a test result from my doctor. But it was getting close to the departure date and I was anxious that the airfare might jump up. So what did United Air offer me in this situation? A fare lock. I paid $7.99 to have a 7-day option to purchase the ticket at the prevailing price on that day regardless of how the price might fluctuate over the following week. For a 72-hour lock, they charged $5.99. As I was first writing this lecture note in August 2015, just out of curiosity, I checked the price again. The 7-day lock cost $11.99. So I inferred that airfare must have turned more volatile from May to August 2015. As a matter of fact, I was tempted to write a code to automatically collect the fare lock price once a day so as to back out the pricing model. Is the society better because of this product? At least I was able to wait for my test result without the added anxiety about airfare fluctuations.

Sure, this is not a financial product but the underlying message is the same. In the presence of risk, it helps if you could give people more flexibility in the kind of risk they take. This example is simpler and easier to communicate. But financial products serve the same purpose for individuals and corporations. So why are the general public so negative about financial innovation? The former Fed Chairman Paul Volcker was quoted in saying that the most important financial innovation that he has seen in the past 20 years is the automatic teller machine. He then added that this is more of a mechanical innovation than a financial one. The practices of some financial institutions and individuals deserve 100% of the criticism. There is not question about it. But, in my personal opinion, the criticism piled up on the innovation itself is perhaps misplaced.

# 4 Financial Markets

I grew up in Shanghai with very little knowledge about stocks or bonds. It's possible that these names had never even showed up in my vocabulary. At that time in China, people read Philosophy books and looked up to scientists. My dream was to become Marie Curie. Dealing with money was just beneath me. In November 1990, several months after I left Shanghai for the US, the Shanghai Stock Exchange was re-established. The rest, as they say, is history. It is probably not an exaggeration to say that a one-day coverage of China in the *Wall Street Journal* today equals that of a year back in 1990. In any case, when I finally concluded that Physics and myself had no future together, I was already in New York. A friend told me about Finance and recommended me to read Bernstein's book. For someone with absolutely no knowledge about Finance, it was truly an eye-opener. Let me borrow his words in describing the financial markets:

> Financial markets are among the most dazzling creations of the modern world. Popular histories of financial markets from the City of London to Wall Street tell the story of panics, robber barons, crooks, and rags-to-riches tycoons. But such colorful tales give little hint of the seriousness of the business that goes on in those markets. John Maynard Keynes once remarked that the stock market is little more than a beauty contest and a curse to capitalism. And yet no nation that has abandoned socialism for capitalism considers the job complete until it has a functioning financial market.
>
> Simply put, Wall Street shapes Main Street. It transforms factories, department stores, banking assets, film producers, machinery, soft-drink bottlers, and power lines into something that can be easily convertible into money and into vehicles for diversifying risks. It converts such entities into assets that you can trade with anonymous buyers or sellers. It makes hard assets liquid, and it puts a price on those assets that promises that they will be put to their most productive uses.
>
> Wall Street also changes the character of the assets themselves. It has never been a place where people merely exchange money for stocks, bonds, and mortgages. Wall Street is a focal point where individuals, businesses, and even entire economies anticipate the future. The daily movements of security prices reveal how confident people are in their expectations, what time horizons they envisage, and what hopes and fears they are communicating to one another.[4]

[4]Introduction of "Capital Ideas," by Peter Bernstein.

Over the next few months, I hope to be able to teach financial markets with this sense of awe, which I felt when I first learned about them. At a personal level, I am not too involved with the markets. Whatever money I have, I put them in index funds and wish to see them again after retirement in a few years. Buy-and-forget pretty much sums up my investment strategy. And yet, I follow the development of the markets with great interest. I view it as a stage with great drama. There are uncertainty, human emotion, and, surprisingly, logic, rationality and regularity. In this course, we will cover the three key markets: equity, fixed-income, and derivatives. More specifically, we will study in depth the US equity markets, the equity index options, US Treasury bonds, corporate bonds, interest-rate swaps, and credit derivatives. If time permits, we will also cover currency. In the next section, I'll explain the topics to be covered in more detail.

# 5 Topics to be Covered

As an academic discipline, what Finance can offer to financial markets is a "cool head." A roller-coaster ride might be exciting initially, but after a few ups and downs, anyone with a sensible mind would ask, is there more to it? Likewise, when being buried deeply in the ups and downs of financial markets, most of us welcome the opportunity to extricate ourselves from the noise and busyness to get a better view. We are often told to learn from our own experiences in life. Empirical Asset Pricing is about learning from the experiences in the financial markets. For this, we have a list of quantitative tools and models at our disposal, which have been developed and widely used. By now, these tools and models have become part of the language on Wall Street. Warren Buffett might have the luxury of not knowing them or even making fun of them. But before you become him, you probably need to know. In any case, these are fun tools and models to learn, especially when you apply them to real financial data.

Let me list the topics to be covered over this semester. I've put a class number to each topic so as to have some discipline. But once in a while I might have to modify our schedule. By the end of the semester, if we somehow end up with extra time, I hope to be able to cover market micro-structure with topics like price discovery, information trading, market making, and high-frequency trading.

1. **Introduction: Class 1**.
2. **Equity:**
   (a) **Class 2**. Alpha and Beta.

(b) **Classes 3 & 4**. Equity in the Cross-Section, Fama-French Three-Factor Model.

(c) **Classes 5 & 6**. Other Cross-Sectional Trading Strategies and Currency Carry Trades.

(d) **Class 7**. Equity in the Time-Series, Time-Varying Expected Returns.

(e) **Classes 8 & 9**. Equity in the Time-Series, Time-Varying Volatility.

3. **Option:**

(a) **Class 10**. Option: Introduction and the Black-Scholes Model.

(b) **Class 11**. **MidTerm Exam.** (covers Classes 1 to 9.)

(c) **Class 12 & 13**. Option: Model to Data, Volatility Smirks and Tail Risk.

(d) **Class 14**. Option: Beyond the Black-Scholes Model.

4. **Special Topic:**

(a) **Classes 15 & 16**. Risk Management.

5. **Fixed Income:**

(a) **Class 17**. Bond: Yield and Duration.

(b) **Class 18**. Bond: Yield Curve.

(c) **Class 19**. Bond: Term Structure Models.

(d) **Class 20**. Bond: Interest Rate Swaps.

(e) **Class 21**. Credit: Corporate Bonds and the Merton Model.

(f) **Class 22**. Credit: Credit Default Swaps and Other Models of Default.

6. **Portfolio Management:**

(a) **Class 23**. The Process of Portfolio Management and Optimal Risky Portfolio.

(b) **Class 24**. The Black-Litterman Asset Allocation Model.

# 6 Quantifying Risk

A professor went to a Zen master, inquiring about Zen. The master invited him to sit down and have a cup of tea. When the tea kettle arrived, the master started to pour hot water into the cup. Soon, the cup was full, but the master kept on pouring. The professor exclaimed, "Master, master, the cup is full. It cannot take any more water!" The master responded, "That's right, Professor. Like this cup, your mind is full of your own views and opinions. How can I teach you Zen?"

So whatever opinion and knowledge you might have about what I am going to teach, please throw them out and keep your mind open. You might be able to do a histogram with eyes closed. You might know the pdf and cdf of a normal distribution inside out. Still, your understanding might be lacking. Going through the motions is easy. Understanding the insight takes more attention, patience, and reflection. Just like anything in life, it is the insight that really matters. Knowledge without insight is dead knowledge.

## 6.1 Data

To understand what uncertainty means, let's start with the past experiences of uncertainty in the US stock market. The time-series plotted in Figure 1 contains the annual stock returns in the US markets from 1927 through 2015. It uses the CRSP value-weighted index, which includes all stocks traded on the three major US exchanges (NYSE, AMEX, and NASDAQ). It is an index preferred by academics. The reported returns are calculated from year-end to year-end, including both capital gains and distributions (i.e., dividends). Another index that would work equally well to represent the overall market is the S&P 500 index. In other words, from a value-weighted perspective, the entire stock market can be very well captured by those 500 large-cap stocks included in the S&P 500 index.

Later in this class, we will sample the data at a daily frequency, and things will look very different. I am using the annual sampling frequency as an example here because it is a widely used horizon. (Also, it keeps our example simple with relatively small amount of data points.) In assessing the performance of a stock or a managed portfolio, the average annual return is often the first benchmark. As such, you should be very comfortable with comparing performance numbers at this frequency.

As a first step, study the plot, follow the ups and downs, absorb the information at an intuitive level. Be curious, and ask yourself questions. This is your random walk down wall street. For example, as an exam question, I could ask you, in the history of the US stock market, what were the worst one-year returns? When did they happen? I might not expect

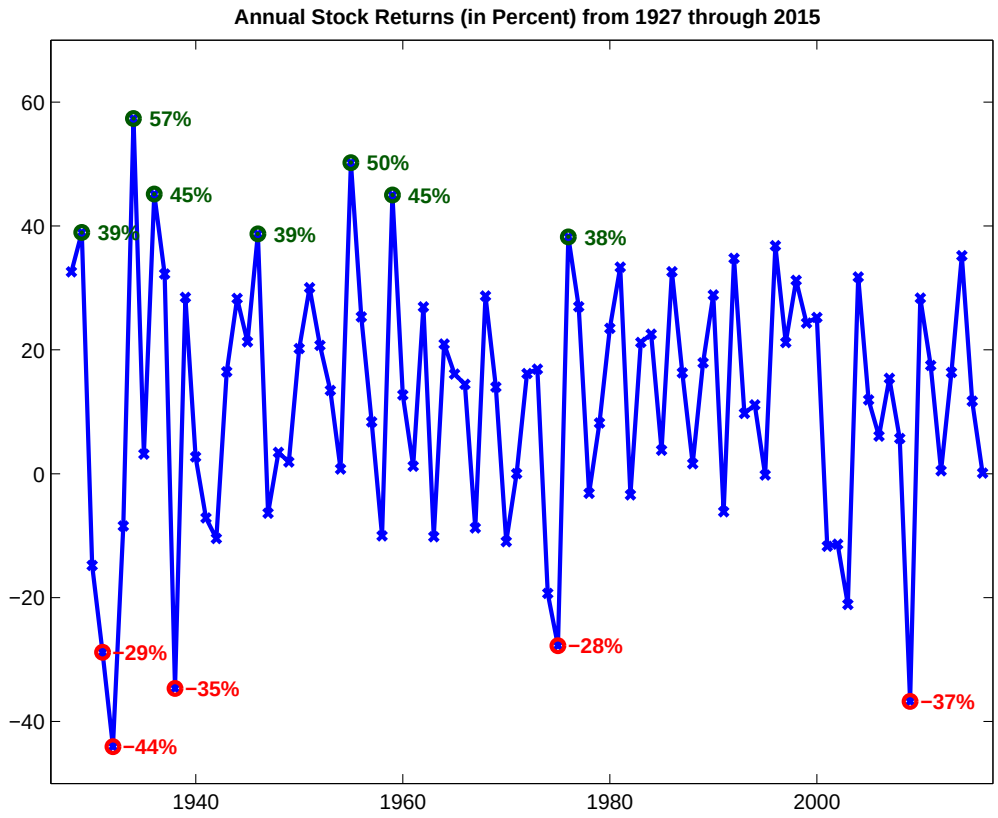


Figure 1: Time-series of annual stock market returns from 1927 through 2015. Returns are calculated using the CRSP-value weighted index, which includes all stocks traded on the US exchanges. Source: Prof. Ken French's website.

you to be so accurate as to tell me -44% in 1931 or -37% in 2008. But I would expect you to know that it was closer to -40% than -20% or -60%. I also would expect you to know about the great depression in the 1930s and the more recent crisis in 2008. Moreover, if I ask you, do you consider a 20% annual return to be normal in the US stock market? You should be ready to articulate your response with the help of the time-series data. Or suppose I tell you that the Shanghai Stock Exchange (SSE) composite index was at 2,115.98 at the end of 2013 and increased to 3,234.68 at end end of 2014. Using the US experiences, how unusual is such an annual performance? These are not hard questions once you really get yourself into the data.

Don't force yourself to memorize these numbers. You will soon forget. What is the point? You need to be interested. You don't need to force yourself to memorize your hometown, your good friends, or your family, do you? I once read someone describing how it was like working with Robert Rubin, the former Treasury Secretary and a long-time executive in Goldman Sachs. I don't remember the exact description but the impression I had was that, when presented with data and plots, this guy would just dive into them with his pencil and mind. Suppose you are presented with a plot like Figure 1 for the first time. And you react with folded arms and the look of "what is the big deal" and "please impress me," then you probably are not going to enjoy Finance as a profession.

Finally, if we look more closely, we will also notice that the large movements are related to the economic conditions. Using the business cycles dated by the NBER, we can see that the largest five negative returns all happened during recessions. Both 1931 (-44%) and 1930 (-29%) fell in the middle of the severe recession from August 1929 to March 1933; 2008 (-37%) was in the recession from December 2007 to June 2009; 1937 (-35%) in the recession from May 1937 to June 1938; and 1974 (-28%) in the recession from November 1973 to March 1975.[5] On paper, these events might be distant and without any real connection. But if you had to live through any of these crises, the impact would be totally different. For example, Warren Buffett was born in August 1930, ten months after the great crash of 1929 and right in the midst of the great depression. His dad, being a stock broker for the Union State Bank, was having trouble feeding the family.[6] Clearly, an upbringing like this helps shape one's investment philosophy and, more broadly, one's attitude toward life. A trader/investor who has lived through a crisis like 2008 would look upon the financial markets with a new set of eyes. So if you would like to become a great investor, try to make a closer connection to the historical events, through books, newspapers, and accounts by older and wiser people. Make these events real for yourself and they would be a treasure sitting in the background,

[5]For a complete list of the NBER business cycles, please go to `http://www.nber.org/cycles.html`.

[6]See *"The Snowball: Warren Buffett and the Business of Life,"* by Alice Schroeder.

emerging when the moment calls for them. After all, life is an endless cycle of ups and downs. So are the financial markets.

## 6.2 Histogram

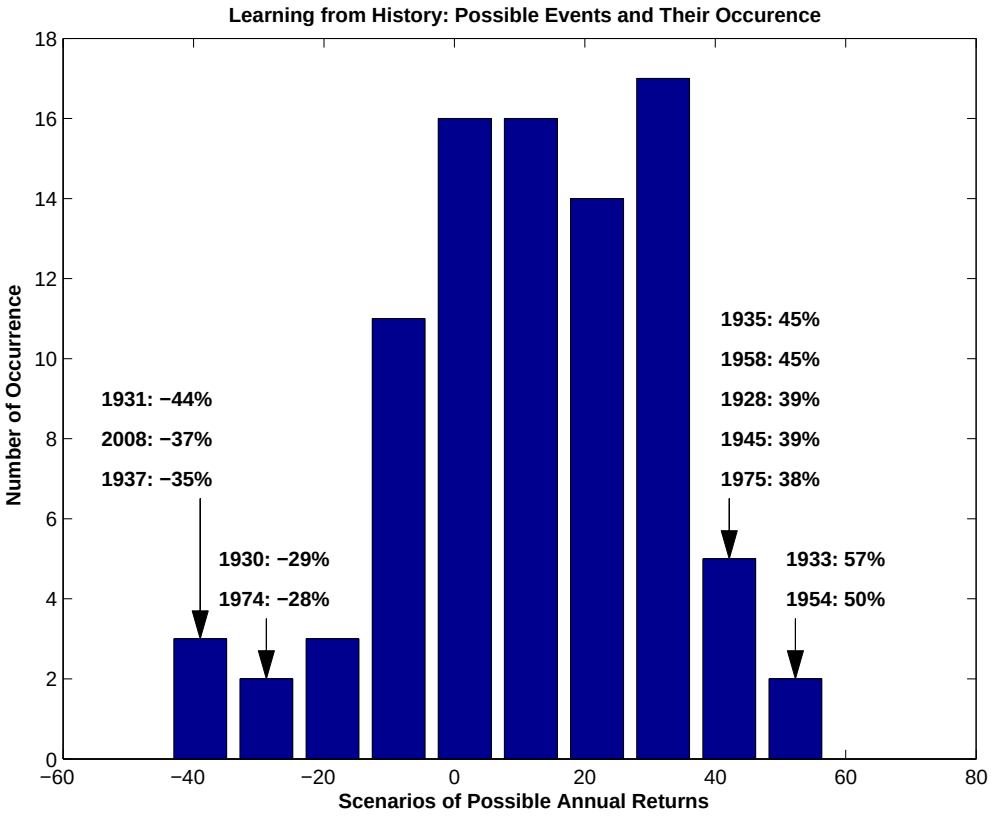


Figure 2: Histogram of stock market returns from 1927 through 2015.

A histogram is like a CT scan on financial returns. In a CT scan, all parts of an image are captured at the same time. In a histogram, however, the image is a compilation of data points collected over a long time span — in some cases, many decades. This is because the financial markets, especially the equity markets, are known to be volatile. So with a few years of data, all you see is noise; but with years of repetition, the noise gets washes out and the valuable signal emerges with more precision. So doing a histogram is like pointing a camera at the stock market for a really really long time. In our current example, the camera has been on for 89 years from 1927 to 2015.

Comparing Figures 1 and 2, you notice that a histogram transforms the dynamic time-series data into a static one. In creating the histogram in Figure 2, we do not care at all

about the sequence of the events – which data point happens first and which one happens later. All we need are the outcomes: the 89 data points, i.e., 89 returns. We line up the returns from the worst (-44.04%) to the best (57.35%), chop the entire interval evenly into N bins (N=10 in our example), with each bin the size of (0.5735+0.4404)/10=0.1014. We then count how many data points (i.e., returns) fall into each bin. So that's a histogram. Sounds really easy, and it is indeed very easy. What is the big deal?

Conceptually, the transformation from Figure 1 to Figure 2 is a significant one. It changes how we view the data. In Figure 1, you are bombarded with uncertainty. In Figure 2, we step out of it and begin to deal with it, by developing a *regularity* about the uncertainty. In order to do so, however, we have to make some non-trivial assumptions. In creating the histogram in Figure 2, the underlying assumption is that every year from 1927 to 2015, the uncertainty in the stock market is exactly the same. In other words, sitting in the background is a distribution machine (like the wizard of oz), which spits out returns with a regularity (i.e., likelihood) over a certain range. Every year, this distribution gives you one realization. You record the outcome. One year later, it totally forgets what it gave you the year before and draws from that identical distribution one more time and give you a new realization. You dutifully record the outcome. After 89 years of repetition, you tell yourself, I should be smart about it. After all, it is going to draw from that same distribution machine again. Let me use the 89 data points I have recorded so far to paint a picture of that distribution. Out comes the histogram in Figure 2. Now, instead of swimming in the sea of uncertainty like in Figure 1, you are looking at the uncertainty from the lens of Figure 2. The future is still uncertain, but you are armed with a tool to deal with the future uncertainty. You know what to expect.

Of course, this is only true if the world functions in this forgetful, and yet consistent way. In Statistics, this underlying assumption is called *iid*: stock returns are independent (forgetful) and identically (consistent) distributed. It is exactly because of this *iid* assumption, we are willing to throw away the sequencing information in doing the histogram and focus only on the outcomes. Otherwise, the analysis will be done differently. Suppose that the real world is not *iid*. Instead, one year of good performance is more likely to be followed by another year of good performance. Then for sure, we will not be throwing away the sequencing information. Later in the semester, we will discuss this possibility of predictive returns. Throughout the semester, you will notice that in Finance we often make strong assumptions first, and examine the markets under these assumptions. Then we realize that perhaps the initial assumptions are too strong. We then relax the assumptions and re-examine the markets. This is the typical process of an empirical investigation. Strong assumptions are

never the problem, but making financial decisions under some assumptions and yet not being aware of them is often the problem.

At this point, let me summarize the things we know and the things we are not sure about. Let's suppose that now is year $t$ and let's use $\tilde{R}_{t+1}$ to denote the stock returns over the next year. Notice that $\tilde{R}_{t+1}$ a random variable. I put a tilde on top of it to remind you that it is not a number. Associated with this random variable is the histogram in Figure 2, which tells you all of the possible scenarios and their likelihood. This is what we know about about $\tilde{R}_{t+1}$: the future outcome will be drawn from a distribution centered around a value, which we call the expected stock return $E(\tilde{R}_{t+1})$. Taking the average of the 89 data points, we say that $E(\tilde{R}_{t+1})$ can be approximated by the historical average of 12%. The eventual outcome of $\tilde{R}_{t+1}$ remains uncertain. A year later, we will get to see its realization. It could be something like 2008 (-37%), or it could be something like 1954 (50%). *Ex ante* (i.e., before the fact), we can tell you the probability of such outcomes. But *ex post*, what will eventually happen? We are not sure. This is the uncertainty faced by everyone, no matter how powerful that person might be. Maybe it is to express the frustration over the lack of knowing, we call this uncertainty *risk*.

The limitation of such an empirical exercise of learning from the history is the history itself. If a certain kind of risk has not yet happened, then it will not be part of our histogram. For example, before the S&P 500 index dropped by -20% over just one day on October 19,1987, this kind of one-day event was not in anyone's histogram. Now that it happened, it adds to the left tail of the daily distribution and our "model" is updated.

## 6.3 Probability Density Function (PDF)

The histogram in Figure 2 tells us how often an event falls within a certain bin. It is not a probability distribution yet because a probability distribution needs to add up to one. From Figure 2 to Figure 3, the shape of the distribution remains exactly the same, but the labeling of the y-axis has changed from "number of occurrences" to "probability density." If you go over the Matlab code attached in the end of this note, you will see that I scaled the entire plot by a constant: $0.1014 \times 89$, where 0.1014 is the width of the bin and 89 is the total number of events. This way, the entire area of the blue bars sums up to one. The probability of all likely events adds up to one. A histogram now becomes a probability density function. We call it an empirical pdf because it builds from the data. Now let's introduce a model to mimic this empirical distribution.

Carrying around an empirical distribution is cumbersome. The next step is to introduce a well behaved analytic distribution to approximate it. This is where normal distribution

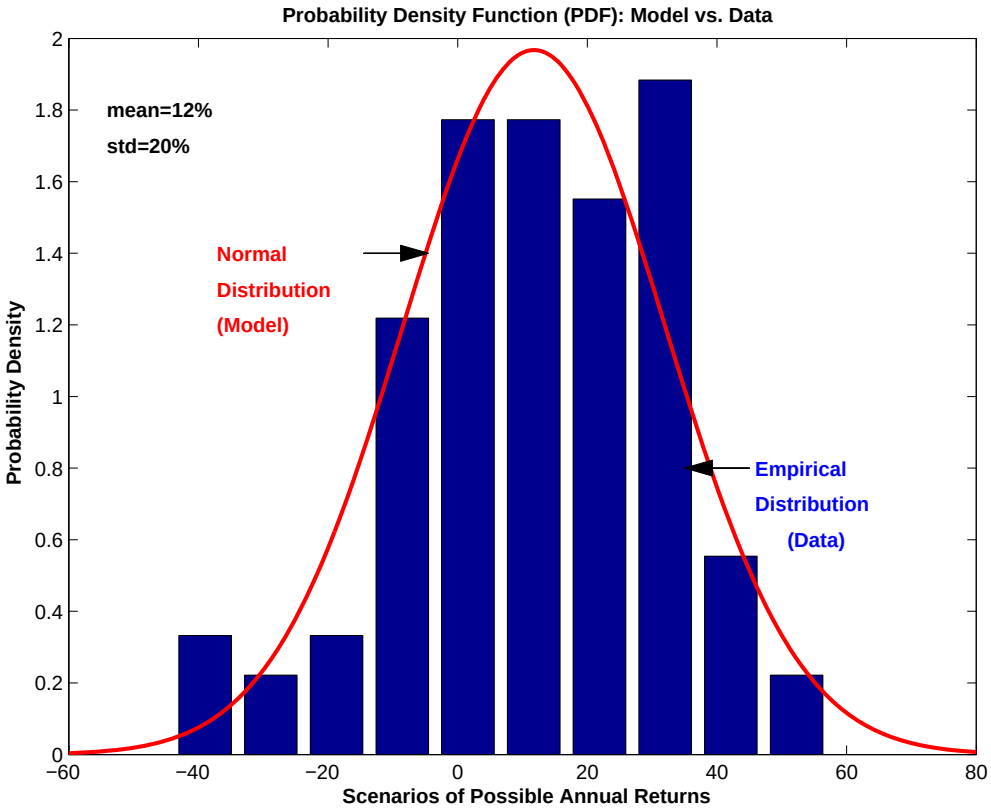


Figure 3: Probability density function, model vs. data.

comes in. You will see that normal distribution plays a very important role in Finance. Let's start with the probability density function of a normal distribution with mean $\mu$ and volatility $\sigma$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

If this looks discouraging, don't be. Because you can do `normpdf(x,`$\mu$`,`$\sigma$`)` in Matlab and out comes the value of $f(x)$. In Excel, there is something that is just as easy. Somehow, computers make math more accessible. Those people who are good at equations suddenly become less attractive. In any case, this is how I plotted the red line in Figure 3. I first calculated the mean (approximately 12%) and standard deviation (approximately 20%) of my 89 data points. Then, for each value in the of stock returns, say -44%, I plugged in `normpdf(-0.44,0.12,0.20)` to get its pdf value, without having to remind myself the exact expression of a Gaussian function.[7]

Let's now formalize what we have done so far. Out of the historical experiences of 89 returns, we are able to develop the following regularity about the risk in the stock market. Let $\tilde{R}_{t+1}$ be the uncertain return over the next year. Learning from the history, we know that its expected value is $\mu = 12\%$ and its standard deviation (i.e., volatility) is $\sigma = 20\%$. Moreover, we know that its distribution can be approximated by $f(x)$. All of this can be summarized by the following model of stock returns:

$$\tilde{R}_{t+1} = \mu + \sigma\, \tilde{\epsilon}_{t+1}$$

where $\tilde{\epsilon}_{t+1}$ is a standard normal distribution: $E(\epsilon) = 0$ and $\text{std}(\epsilon) = 1$.

We do not have time to formally test the model, but from Figure 3, it seems to be a reasonable approximation of the data. Not perfect, but reasonable and useful, as you will see. But as useful as a model might be, you should always use it with extreme caution. The limitation of a model is two fold. First, it builds from the empirical distribution, which itself is limited by our experiences. Second, it is an analytic approximation of the much richer reality. Some of the financial crises happened exactly when investors become too comfortable with their models in their pricing, hedging, and trading.

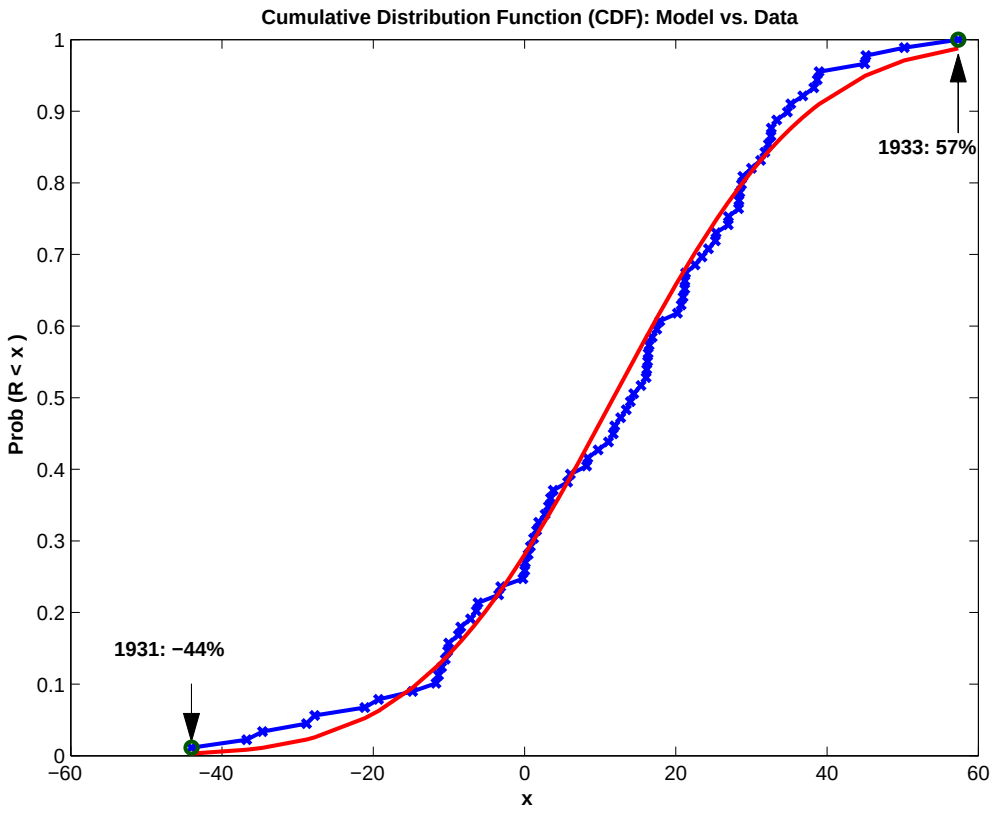


Figure 4: Cumulative distribution function, model vs. data.

## 6.4 Cumulative Distribution Function (CDF)

In order to plot the blue line in Figure 4, I line up the 89 data points from left to right, from the most negative to the most positive. So -44% in 1931 is my leftmost data point, with -37% in 2008 standing next to it, and 57% in 1933 is my rightmost data point. There are 89 points in total. So the values of the empirical CDF are:

$$\begin{aligned} \text{Prob}(\tilde{R}_{t+1} \leq -0.44) &= 1/89 \\ \text{Prob}(\tilde{R}_{t+1} \leq -0.37) &= 2/89 \\ &\ldots \\ \text{Prob}(\tilde{R}_{t+1} \leq 0.57) &= 89/89 \end{aligned}$$

That is how the blue line is constructed. Isn't it simple? Now we have an empirical CDF.

[7] To be more precise, I use `normpdf(-0.44,mean(RM/100),std(RM/100)`, where RM is the time-series of the 89 market returns in percentage and mean and std are the Matlab functions to calculate mean and standard deviation. See the Matlab code in the Appendix.

The CDF for the model is just as simple. Mathematically, for any value $x$, it is

$$\text{Prob}\left(\tilde{R}_{t+1} \leq x\right) = \int_{-\infty}^{x} f(z)\, dz\,,$$

which might look intimidating until you realize that you can use Matlab and do:

$$\text{Prob}\left(\tilde{R}_{t+1} \leq x\right) = \texttt{normcdf}(\texttt{x}, \mu, \sigma)\,.$$

This is how I plotted the red line, I plugged all 89 returns into `normcdf` with $\mu = 12\%$ and $\sigma = 20\%$.[8] That is why the red line is not as smooth as a textbook example: the 89 data points are not evenly spaced. For example, to move from -44% to -37% is a pretty large gap. In my program, I do normcdf on these two returns and ask Matlab to link the two points with a straight line. To get a smoother line, I could have asked the normcdf to evaluate many more returns in between -44% and -37% and then connect the points. This is the advantage of a model over data. The data only gives us two experiences, -44% and -37%, and nothing in between. But the model can extrapolate to places that experiences did not take us. Needless to day, this is always the danger of a model, especially when we mistake the model as the reality.

## 6.5 Models are Limited

The normal distribution model is a corner stone of Finance. In the CAPM, we care only about the first two moments (mean and variance) of stock returns. Implicitly, we are treating returns as a normal distribution with $\mu$ and $\sigma$. In the Black-Scholes model, this assumption of normal distribution is explicitly given. That is why in the pricing formula, you see normcdf here and there. This model offers simplicity and elegance, a great vehicle to carry us far. But once in a while, it carries us too far.

Let's take a look at an example when this model fails. Figure 5 plots the daily returns of the S&P 500 index. Comparing with the time-series in Figure 1, we have more data points and the returns fluctuate within a much narrower band. Over an year, the volatility is about 20%. Over a day, it is about 1%.

Using the daily data, Figure 6 plots the CDF for the left and right tails. The red line is the CDF produced by normal distribution, with the same mean and variance as the empirical distribution, and the blue and green lines are the CDFs produced by the historical

[8] Again, for illustration purpose, I am using 12% and 20% as the sample mean and standard deviation. In practice, the actual sample mean and standard deviation for the 89 data points are used.

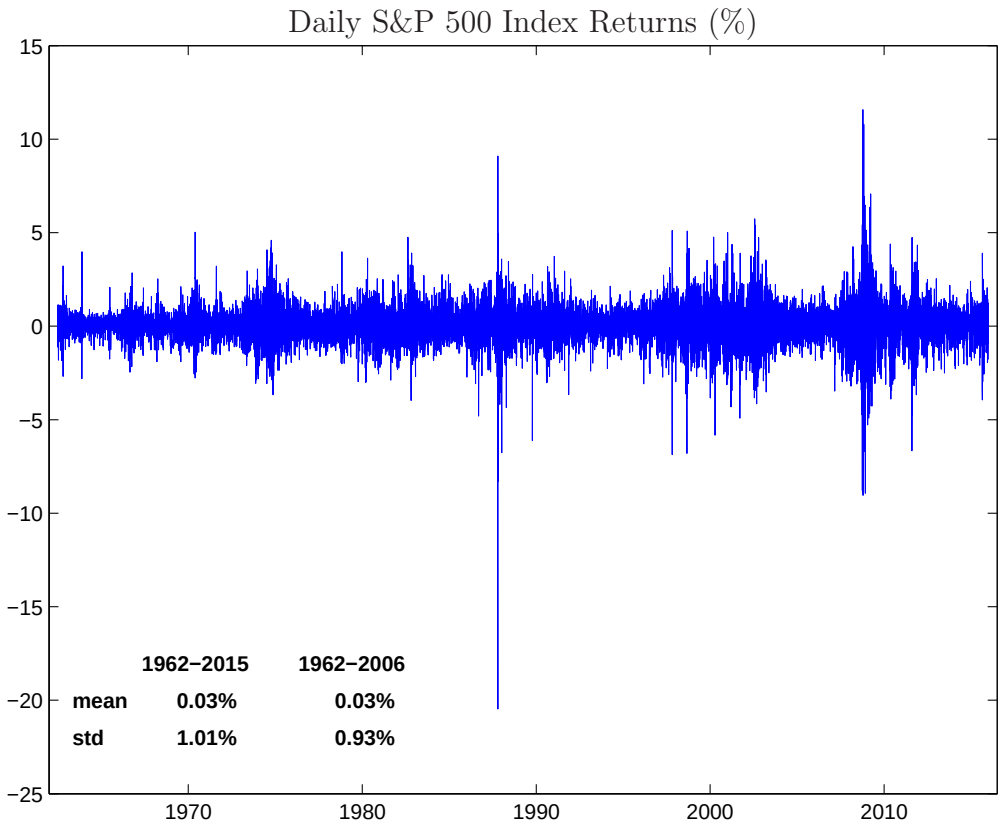


Figure 5: Time-series of daily returns on the S&P 500 index from 1962 through 2010. Source: Wharton Research Data Service (WRDS).

experiences. As it is evident in the plots, both the left and right tails are much thicker in the data. If your financial instruments are very sensitive to the tails (e.g., a far out-of-the-money put option), then this difference would really matter. In fact, this is one of the first places where the Black-Scholes model fails to perform.

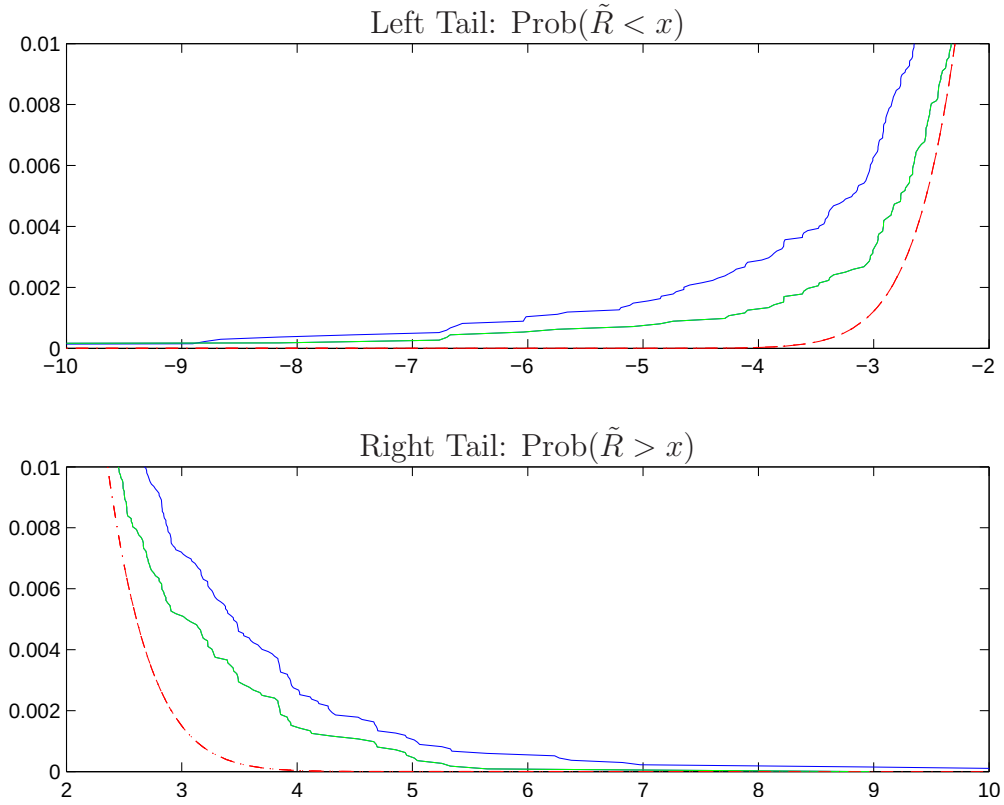


Figure 6: Cumulative distribution function using daily returns, model vs. data. The x-axis measures the daily standard deviation moves: daily returns normalized by the sample standard deviation. The y-axis measures the cumulative probability: 0.01 is 1%. The blue line is the empirical CDF generated by daily returns from 1962 through 2015. The green line is the empirical CDF generated by daily returns from 1962 through 2006. The red line is the analytic CDF generated by a normal distribution with the same mean and standard deviation as the entire sample (1962-2015).

# A Matlab Code

Code 1: Histogram.m

```
load FF_Factors_Annual.txt; % Posted on Stellar under Data. Orignally
    from Prof. Ken French's website.
Data=FF_Factors_Annual; %Data Content: year, Mkt-RF, SMB, HML, RF

YR=Data(:,1); RF=Data(:,end); RM=Data(:,2)+RF; %RM=Mkt, RF=riskfree
Time=datenum([YR 12*ones(length(YR),1) 31*ones(length(YR),1)]); %I always
    prefer to work with the Matlab time. All returns are realized
    at the year end.

% Plot the time-series data
figure(1); clf;
h=plot(Time,RM,'bx-','LineWidth',2);
datetick('x','yyyy');
title('\bf Annual Stock Returns (in Percent) from 1927 through 2015');
hold on;

% List the worst five years
[tmp,Index]=sort(RM); Worst=[]; Best=[];
for i=1:5,
  Worst=[Worst; [year(Time(Index(i))),round(RM(Index(i)))]];
  p=text(Time(Index(i))*1.0005,RM(Index(i)),['\bf' num2str(round(RM(Index(i)
      ))) '%']);
  set(p,'Color','red');
  plot(Time(Index(i)),RM(Index(i)),'ro','LineWidth',2);
end
% List the best seven years
for i=0:6,
  Best=[Best; [year(Time(Index(end-i))),round(RM(Index(end-i)))]];
  p=text(Time(Index(end-i))*1.0005,RM(Index(end-i)),['\bf ' num2str(round(RM
      (Index(end-i)))) '%']);
  darkgreen=[0 0.35 0];
  set(p,'Color',darkgreen);
  plot(Time(Index(end-i)),RM(Index(end-i)),'Marker','o','Color',darkgreen,'
      LineWidth',2);
end
```

```
hold off;
axis([datenum([1925 12 31]) datenum([2016 12 31]) -50 70]);

% Histogram
figure(2)
[Occurence Scenario]=hist(RM);
bar(Scenario,Occurence);
axis([-60 80 0 18])
xlabel('\bf Scenarios of Possible Annual Returns');
ylabel('\bf Number of Occurrence');
title('\bf Learning from History: Possible Events and Their Occurence');
hold on;

% Mark the left tail
text(-55,9,['\bf ' num2str(Worst(1,1)) ': ' num2str(Worst(1,2)) '%']);
text(-55,8,['\bf ' num2str(Worst(2,1)) ': ' num2str(Worst(2,2)) '%']);
text(-55,7,['\bf ' num2str(Worst(3,1)) ': ' num2str(Worst(3,2)) '%']);;
arrow([Scenario(1) 6.5],[Scenario(1) 3]);
text(-37,5,['\bf ' num2str(Worst(4,1)) ': ' num2str(Worst(4,2)) '%']);
text(-37,4,['\bf ' num2str(Worst(5,1)) ': ' num2str(Worst(5,2)) '%']);
arrow([Scenario(2) 3.5],[Scenario(2) 2]);

% Mark the right tail
text(50,5,['\bf ' num2str(Best(1,1)) ': ' num2str(Best(1,2)) '%']);
text(50,4,['\bf ' num2str(Best(2,1)) ': ' num2str(Best(2,2)) '%']);
arrow([Scenario(end) 3.5],[Scenario(end) 2]);
text(40,11,['\bf ' num2str(Best(3,1)) ': ' num2str(Best(3,2)) '%']);
text(40,10,['\bf ' num2str(Best(4,1)) ': ' num2str(Best(4,2)) '%']);
text(40,9,['\bf ' num2str(Best(5,1)) ': ' num2str(Best(5,2)) '%']);
text(40,8,['\bf ' num2str(Best(6,1)) ': ' num2str(Best(6,2)) '%']);
text(40,7,['\bf ' num2str(Best(7,1)) ': ' num2str(Best(7,2)) '%']);
arrow([Scenario(end-1) 6.5],[Scenario(end-1) 5]);
hold off

% Plot the distribution, pdf
figure(3); clf;
D_X=mean(diff(Scenario/100)); %Scenario and RM are in return space. To
    be careful, I always do the math in decimal.
```

```
N_norm=D_X*sum(Occurence);
bar(Scenario,Occurence/N_norm); %rescale the number of occurence to get
    probability density.
hold on;
PDF=normpdf(-0.60:0.01:0.80,mean(RM/100),std(RM/100)); %probability
   density function.
h=plot(-60:1:80,PDF,'r-','LineWidth',2);
hold off
V=axis; axis([-60 80 V(3:4)]);
text(-38,1.4,'\bf Normal','Color','r');
text(-38,1.3,'\bf Distribution','Color','r');
text(-38,1.2,'\bf (Model)','Color','r')
arrow([-14.5 1.4],[-4.5 1.4]);
text(45,0.8,'\bf Empirical','Color','b');
text(45,0.7,'\bf Distribution','Color','b');
text(50,0.6,'\bf (Data)','Color','b')
arrow([45 0.8],[35 0.8]);
text(-55,1.8,['\bf mean=' num2str(round(mean(RM))) '%']);
text(-55,1.7,['\bf std=' num2str(round(std(RM))) '%']);
xlabel('\bf Scenarios of Possible Annual Returns');
ylabel('\bf Probability Density');
title('\bf Probability Density Function (PDF): Model vs. Data');

% Plot the distribution, cdf
figure(4); clf;
mean_R=mean(RM/100); std_R=std(RM/100); %to be safe, always do my math
   in decimal.
sorted_RM=sort(RM); Y_vec=cumsum(ones(length(RM),1))/length(RM);
Y_norm=normcdf((sorted_RM/100-mean_R)/std_R); %again, do my math in
   decimal.
plot(sorted_RM,Y_vec,'b-x',sorted_RM,Y_norm,'r-','LineWidth',2);
hold on
title('\bf Cumulative Distribution Function (CDF): Model vs. Data');
ylabel('\bf Prob (R < x )');
xlabel('\bf x');
% Mark the left tail
text(-55,0.15,['\bf ' num2str(Worst(1,1)) ': ' num2str(Worst(1,2)) '%']);
arrow([sorted_RM(1) 0.1],[sorted_RM(1) 0.02]);
```

```
plot(sorted_RM(1),Y_vec(1),'o','Color',darkgreen,'LineWidth',2);

% Mark the right tail
text(46,0.85,['\bf ' num2str(Best(1,1)) ': ' num2str(Best(1,2)) '%']);
arrow([sorted_RM(end) 0.87],[sorted_RM(end) 0.98]);
plot(sorted_RM(end),Y_vec(end),'o','Color',darkgreen,'LineWidth',2);
hold off;
```

Jun Pan
junpan@mit.edu

253-3083
E62-624

**Class 2: Alpha and Beta**

This Version: September 9, 2016

# 1 Financial Data

Financial markets are places where people trade on their information. Sometimes, in some corners of the world, people bring their hope and wishes, fear and greed to the markets. But overall, when all of these noises get canceled out, the financial market is a powerful information central. No other platform in this world can gather and process information in a more efficient way. What people bring to the markets might be messy, and the people themselves might be messy, but the output is simple and elegant: price and trading volume.

The process is an organic one. Like rivers flowing from the mountain to the sea, not one person can really force the market to go one way or the other for a sustained period of time. The process is not a flawless one. In the recent 2008-09 financial crises, financial prices failed to reflect the real information. Central banks around the globe had to resort to quantitative easings to offer support: cushion the fall and control the flood. But all in all, financial markets are the best place to gather and collect information about the overall economy and individual companies.

Financial data are the direct output of this process. One thing I would like you to develop over the course of this semester is confidence and ease in handling financial data. As a first step, you should know where to get what kind of data.

- **Bloomberg:** For most practitioners, a Bloomberg terminal is the first place to access the real-time information. When I first joined Sloan in 2000, there was just one (or two) Bloomberg terminal sitting in the basement of E52. In recent years, because of the creation of the Master of Finance program, we have significantly increased the number of units at Sloan. All students should take full advantage of this opportunity.

  Since its beginning in 1981, Bloomberg data-and-information terminals have evolved into something that is more complex and powerful than just a data service. They are used by bankers, traders, and money managers for information gathering, trade

communication and execution, and pricing and risk analysis for global products in equities, fixed-income, derivatives, commodities, and foreign exchange. For example, you can find tools for yield curve analyses, as well as pricing functions for interest rate swaps and credit-default swaps on Bloomberg. This kind of convenience could also be worrisome if users are too lazy to do their own price discovery and start to think of the prices generated by a Bloomberg pricing function to be the real price. Just imagine, when you are using a Bloomberg function, everyone else is using it as well. Group think soon prevails.

While there are other data providers such as Thomson Reuters, Bloomberg remains its dominance in part because of the broad user base of its messaging tool. If your client is on Bloomberg, then you will have to be on Bloomberg to reach him. This also creates the situation of "Too Bloomberg to Fail." On April 17, 2015, before the opening of the US markets, there was a computer-network outage on Bloomberg. The blackout, which started shortly after European markets opened, also caused the UK to postpone a scheduled multibillion buyback of government debt. The £3 billion ($4.5 billion) tender was rescheduled for the afternoon.[1]

- **Datastream:** I have to confess that I am not a big fan of Bloomberg terminals. I don't trade. The real-time feature is not attractive for me. I don't have fellow traders or clients to communicate with. Finally, I don't trust the pricing function built by others. But still, for real practitioners, I can certainly see the importance of a Bloomberg terminal.

  I use data provider for teaching and doing research. So when it comes to downloading data, I like to use Datastream, which offers long time-series data with very broad coverage. For example, I used Datastream to download time-series data on interest rate swaps and credit default swaps, which we will use later in the semester. By contrast, downloading the same amount of time-series data from Bloomberg is painful, to say the least.

  Some of the Sloan machines have Datastream installed. You can open a terminal to navigate the system. It offers a wealth of products. To download the time-series data, you can use the Datastream plugin in Excel.

- **CRSP:** The Center for Research in Security Prices (CRSP) is an impressive collective effort done by people at Chicago GSB, now the Booth School. When it comes to data

[1] "Bloomberg Terminals Go Down Globally," *Wall Street Journal*, April 17, 2015.

on US equity returns, CRSP is the gold standard. A lot of effort and care have been put in by these researchers to clean up the raw data from stock exchanges and properly calculate returns.[2] When I use stock return data, I don't trust any other sources.

In addition to stock data, CRSP provides US Treasury data and Mutual Fund data. This is a link to CRSP Manuals and Overviews.

- **WRDS and its component databases:** The Wharton Research Data Services (WRDS) was initially developed to support faculty research at Wharton. Today, it has become an important platform that hosts a wide spectrum of databases, including CRSP. I've applied a class account for us at WRDS. The username is "finmkt" and you can get the password by emailing me or the TAs. I would encourage you to log on to the system to take a look. These are some useful databases hosted by WRDS:
  - **CRSP:** equity, treasury, and mutual fund data.
  - **Compustat:** firm-level fundamentals including income statements, balance sheets, and flow of funds. The most essential data for Corporate Finance and Accounting.
  - **IBES:** historical earnings estimates by analysts, including EPS, revenue, price target, EBITDA and pre-tax profits. Available on both consensus and detailed levels. Also includes buy-hold-sell recommendations by analysts.
  - **Option Metrics:** contains data on equity options and equity index options.
  - **TAQ:** high-frequency transaction and quote data for all stocks listed on US Exchanges.
  - **TRACE:** transaction-level data of US corporate bonds.
- **Prof. Ken French's Website:** Prof. French provides a valuable service to our community by offering a wide range of portfolios and benchmarks on his website. We will use these data quite extensively in our next few classes and in your group assignments.

# 2 Estimating the Expected Stock Returns

- **Computing returns:** For a publicly traded firms, let $P_t$ be its stock price at the end of year $t$, and $D_t$ be the cash dividend paid out during year $t$. The year-$t$ realized

[2] See Chapter 7 of "Capital Ideas" by Perter Bernstein for a detailed account.

return is,

$$R_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} = \frac{P_t - P_{t-1}}{P_{t-1}} + \frac{D_t}{P_{t-1}},$$

which is the sum of capital gain and dividend yield.

Calculating returns is in fact not as simple as you would think. Above is a textbook example, involving only cash dividends. In practice, one has to take care of periodic firm events including splits, reverse splits, stock dividends, rights offerings, spin offs, etc. Dealing with these issues is so tedious that you would want to quit Finance right away. So we are very thankful to CRSP for taking care of all of these firm events and give us a clean and reliable set of stock return data. If you are the curious type or if you would like to apply the CRSP service to another market, say the Chinese stock market (which lacks a professional service such as CRSP), here is a link to CRSP's Data Descriptions Guide.

One year, I asked my TA for 15.433 to calculate for me monthly returns of Berkshire Hathaway (`BRK`). The numbers he gave me looked suspicious because Warren Buffett is known to be a great investor, but my TA told me that BRK alpha is close to zero. So I went over the data myself and found out that there were three months in the 1970s when the price data were missing. Matlab replaces missing data with zero. So my TA caused Warren Buffett to bankrupt three times (i.e., in my TA's spreadsheet, he had -100% returns in three places in the return column), and yet, Mr. Buffett's alpha came out to be close to zero. This is how impressive Mr. Buffett's performance is. After this, I never ask any of my TAs to do my calculations for me.

- **Estimating the expected return:** For any financial instrument, the single most important number is its *expected* return. This is what attracts an investor to that product in the first place. And yet, the expected stock return is the toughest number to measure in Finance.

  Suppose right now we are in year $t$, and let $R_{t+1}$ denote the stock return to be realized next year. In making our investment decision today, one of the key variables is the expected return: $\mu = E(R_{t+1})$ . It is worthwhile to emphasize that $\mu$ is a number, while $R_{t+1}$ is a random variable drawn from a distribution with mean $\mu$ and standard deviation $\sigma$.

  The standard approach in estimating the expected return is to use past returns and

estimate $\mu$ by taking the sample average:

$$\hat{\mu} = \frac{1}{N} \sum_{t=1}^{N} R_t \, .$$

Let's again use the time-series data on annual stock returns, plotted in Figure 1. In this example, we have a time-series of 88 realized returns.

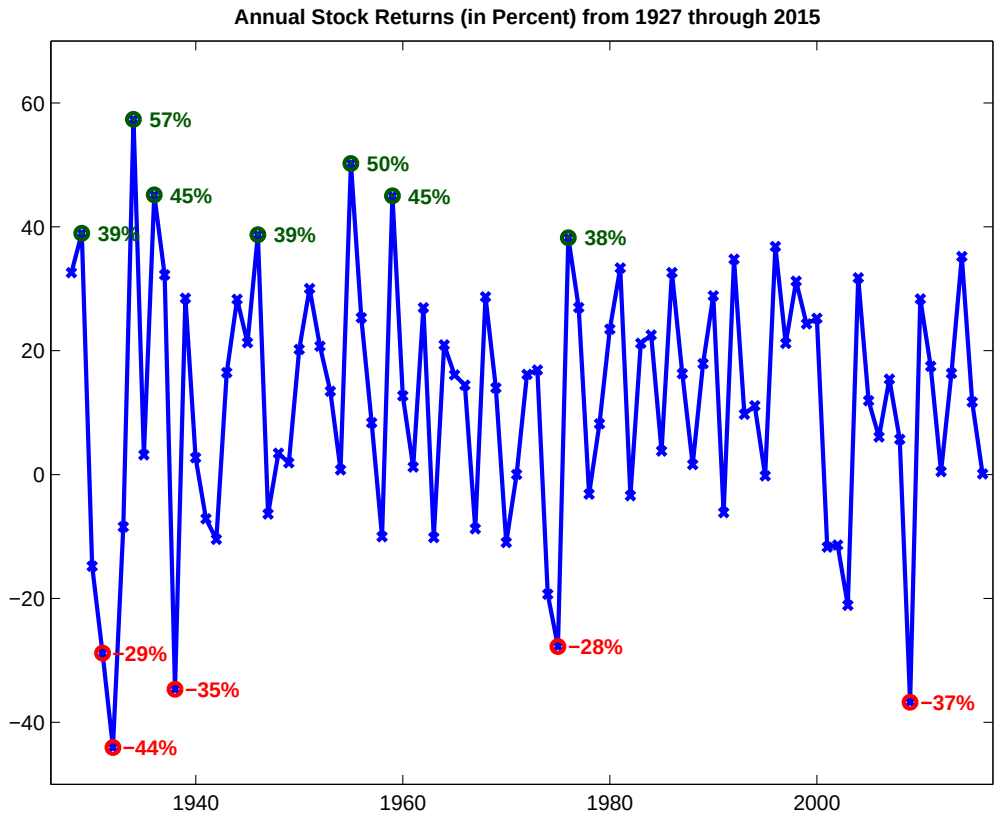


Figure 1: Time-series of annual stock market returns from 1927 through 2015. Returns are calculated using the CRSP-value weighted index, which includes all stocks traded on the US exchanges. Source: Prof. Ken French's website.

One natural question is why can this sample average of past realized returns help us form an expectation of the future? The answer is the same as before. We are assuming that historical repeats itself in such a why that each $R_t$ in the past was drawn from an identical distribution with mean $\mu$ and standard deviation $\sigma$. Moreover, the draw is forgetful in the sense that this year's distribution is independent of past years' distributions. Again, we are using the assumption that stock returns are independent

and identically distributed (*i.i.d.*).

- **Standard Error of $\hat{\mu}$:** At this point, it is important to emphasize the distinction between $\mu$ and $\hat{\mu}$. Specifically, $\hat{\mu}$ is an estimator for an unknown number $\mu$. The estimator $\hat{\mu}$ itself is not a number, it is an average of 88 random variables. As a direct result, $\hat{\mu}$ inherits the noise from the 88 random variables $R_t$:

$$\text{var}(\hat{\mu}) = \text{var}\left(\frac{1}{N}\sum_{t=1}^{N} R_t\right) = \frac{1}{N^2}\sum_{t=1}^{N}\text{var}(R_t) = \frac{1}{N^2}\times N\times\sigma^2 = \frac{1}{N}\,\sigma^2\,.$$

  Notice that the above derivation relies on the *i.i.d.* assumption. In passing through the second equal sign, we use the independent assumption: $\text{cov}(R_t, R_s) = 0$ for $t \neq s$. In passing through the third equal sign, we use the assumption that $R_t$ is identically distributed with a variance of $\sigma^2$.

  Because of the special role of $\hat{\mu}$ as an estimator, we also refer to its standard deviation as the *standard error* of $\hat{\mu}$. To summarize,

$$\text{s.e.}(\hat{\mu}) = \frac{\text{std}(R_t)}{\sqrt{N}} = \frac{\sigma}{\sqrt{N}}\,.$$

  Naturally, an estimator with smaller standard error gives us more precision. From the above equation, the level of volatility in the stock market return plays an important role in determining the noise level of our estimator $\hat{\mu}$. Unfortunately, stock market returns are known to be very "noisy." The only way to improve the precision is by increasing N, the number of observations.

- **Using t-stat:** For the time-series data of annual stock returns from 1927 through 2014, the sample mean is 12% and the sample standard deviation is 20%. So the estimate of $\mu$ is 12% and that of $\sigma$ is 20%. We can now calculate the standard error of this estimator:

$$20\%/\sqrt{88} = 2.13\%\,.$$

  To evaluate how significant an estimate of 12% is in relation to the standard error, we often use the t-stat of the estimator:

$$\text{t-stat} = \frac{12\%}{2.13\%} = 5.63\,.$$

  Effectively, it is a signal-to-noise ratio. The bigger the absolute value of the t-stat, the more significantly away it is from zero. For the rest of the semester, we will use this

rule of thumb: a significant estimate is one whose t-stat has an absolute value greater than 2. In other words, a large value of the estimate itself is not that meaningful. It is only after you scale it with its noise level (i.e., standard error), the measure becomes useful. And our cutoff value is two.[3]

For many emerging markets, the sample averages of their stock returns might be large compared to the US number. But such markets are typically more volatile than the US markets (larger $\sigma$). Moreover, they are also younger (smaller $N$). Factoring in these observations, the t-stat's of $\hat{\mu}$ for these emerging markets are typically much smaller 5.63. For some of the countries (e.g., China), the t-stat's of their $\hat{\mu}$ might not even pass the threshold value of 2, indicating that, statistically speaking, the expected return is not significantly different from zero!

In fact, the current example with a t-stat of 5.63 is as good as it gets when it comes to estimating the expected stock return $\mu$. Nevertheless, the corresponding 95% confidence interval is not that impressive:

$$[12\% - 1.96 \times 2.13\%, 12\% + 1.96 \times 2.13\%] = [7.8\%, 16.2\%]\,.$$

In other words, with 88 years of data and for one of the most stable stock markets in the world, we can only get to this range of 95% confidence interval. That's why I said earlier that the expected stock return is the toughest number to estimate in Finance. As we can see from our analysis, the main reason is the volatility in the stock market.

- **$R_t$ and $\hat{\mu}$ on the same plot:** Just to make a more graphical display of the connection between $\hat{\mu}$ and $R_t$, I plotted in Figure 2 both of their distributions.

  Recall that

  $$\hat{\mu} = \frac{1}{N}\sum_{t=1}^{N} R_t\,.$$

  As a result, the estimator $\hat{\mu}$ is closely connected to stock returns. The signal we really really would like to extract from the realizations of $R_t$ is its mean, i.e., the expected return return. Unfortunately, like everything else in life, it comes as a packaged deal: to get its mean, you also have to take its variance. So the blue line plotted in Figure 2

[3] I'll be happy to explain why the cutoff value is two. But I feel that the explanation will distract us from the Finance content. For those who are interested, the key intuition is that $\hat{\mu}$ is normally distributed (because of the central limit theorem) with a standard deviation of $\sigma/\sqrt{N}$. A cutoff value of 1.96 corresponds to a double-sided test of the null hypothesis that $\mu = 0$ with the significance level of 5%: normcdf(-1.96)*2=0.05. Instead of carrying 1.96, let's just round it to 2.

informs you of the distribution of this packaged deal. Specifically, the blue line is the pdf of a normal distribution with mean 12% and standard deviation of 20%.

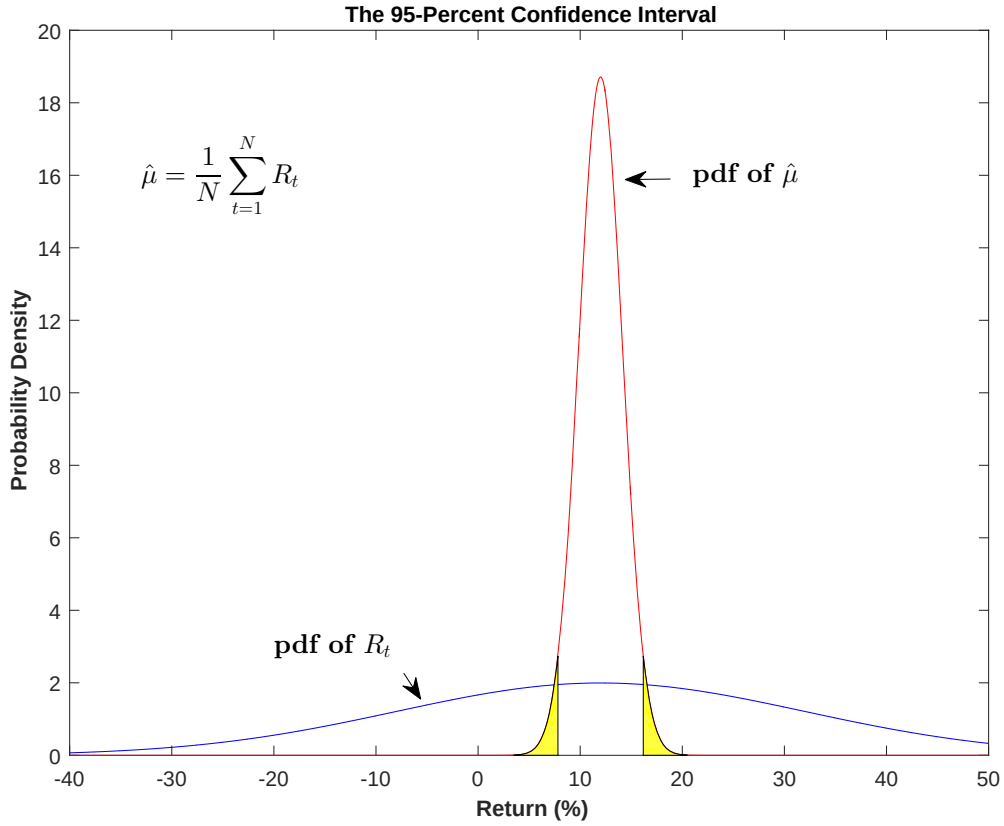


Figure 2: The probability density functions of $R_t$ and $\hat{\mu}$ and the 95-percent confidence interval for the estimator $\hat{\mu}$.

Plotted against the blue line, is the distribution of the extracted signal $\hat{\mu}$. As you can see, with 88 data points, we are able to shrink the noise in the raw data quite significantly. More data, more precision. The exact rate of shrinkage is $\sqrt{N}$. In this plot, I use $N = 88$.

In addition to plotting the pdf of $\hat{\mu}$, I also plotted the 95% confidence interval. For a standard normal distribution, the critical value of a two-sided 5% tail is 1.96. The left and right cutoff values for our current example are $0.12 - 1.96 \times 0.0213$ and $0.12 + 1.96 \times 0.0213$. So the two yellow-shaded areas add up to 5% probability, while the middle unshaded area in the distribution has a total probability of 95%.

In doing a statistical test, the main question we are asking is the following. The estimator is noisy (the standard error is 2.13%). Given this noise level, is the estimated expected return (12%) significantly away from zero? Of course, 12% is different from

zero. Don't be silly. But the question is more about a "fuzzy" 12% like the red distribution in the plot. The red line spells out the level of noise for us. We can now move to the left and reach the cutoff value of the 95% confidence interval and see whether or not this value is still away from zero. Doing a 95% confidence interval is cumbersome. Instead, we calculate the t-stat of the estimator: 12%/2.13% and see how far away it is from 1.96. It has the same effect, but the procedure is simpler. For this class, let's make it even simpler by rounding 1.96 to 2.

- **Why so much emphasis on statistics?** As you can see, I've allocated a significant amount of time on estimation, standard error and t-stat and we've gone through the derivations in quite some details. By doing so, I would like to impress upon you the Statistical foundation of the tools adopted by people in Finance.

  Among the first set of numbers reported in an investment prospectus are the past realized returns of a portfolio manager. There is no mention at all about how noisy these numbers are. Implicitly, that is why a long track record is well respected in the industry. Only when you have a sufficiently long sample, these sample averages become meaningful. Otherwise, they are as good as noise.

  Going forward, we will be working with a variety of estimation results, e.g., regression. I will not ask you to calculate the standard error of a regression coefficient. That, will be too much Statistics in a Finance class. Nevertheless, I would like you to always keep this in mind: as long as you are working with financial data, the numbers you estimate from the data are contaminated by the randomness and noise that is inherent in financial data. Do not treat them as numbers. Treat them as estimators with standard errors and t-stats.

# 3 Estimating Alpha and Beta

- **The risk that matters:** So far, we've focused on one time series and it turns out to be a very important risk factor. According to the CAPM, investors are only rewarded for bearing the *systematic* risk. Any risk that is uncorrelated with this risk should not be rewarded, because it can be diversified away. We are going to use the US aggregate market as a proxy for the systematic risk in the CAPM and take the CAPM model to the data.

  As a starter, let's give it a unique symbol: $R^M$ . Moreover, as it is the convention in this area, alpha's and beta's are estimated using monthly stock returns. So let me

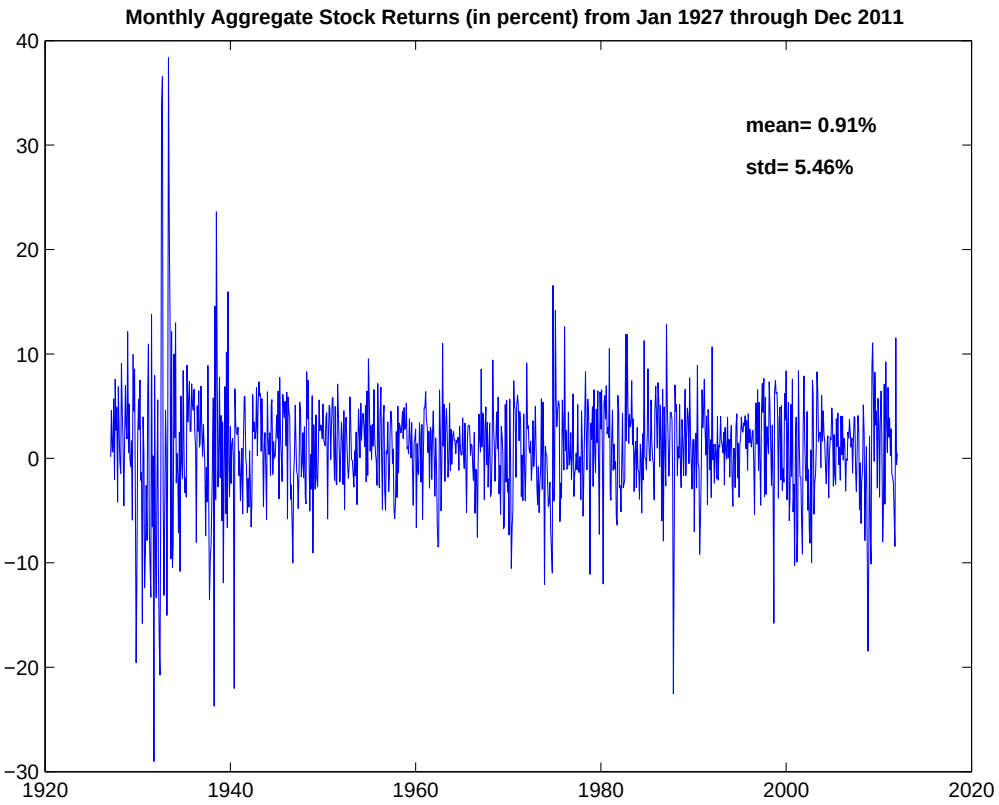


Figure 3: Time-Series of monthly stock returns from 1927 through 2011. Returns are calculated using the CRSP-value weighted index, which includes all stocks traded on the US exchanges. Source: Prof. Ken French's Website.

plot this important time-series using monthly returns in Figure 3, so that you know that the data look like at this frequency. Compared with the annual frequency, the fluctuation is at a smaller range. For this market, the annual volatility is around 20%, while the monthly volatility is around 5.46%. Note that $20\%/\sqrt{12} = 5.77\%$, which is a pretty good approximation if you don't remember the monthly number.

- **The CAPM:** Before we start to use the model quite extensively, let me summarize the key ingredients of the CAPM here.
  - The market risk premium: defined as the expected return of the market portfolio $R^M$ in excess of the riskfree rate $r_f$: $E(R^M) - r_f$. So far, at an annualized level, our estimate for $E(R^M)$ is around 12% and the riskfree rate is on average 4%, making the market risk premium to be around 8%.
  - Beta: The risk of an individual stock, say GE, is measured not by its own volatility,

but by its exposure to the market risk:

$$\beta^{\mathrm{GE}} = \frac{\mathrm{cov}\left(R^{\mathrm{GE}}, R^{M}\right)}{\mathrm{var}\left(R^{M}\right)}$$

– The pricing equation: The reward is proportional to risk:

$$E(R^{GE}) - r_f = \beta^{GE} \times \left(E(R^{M}) - r_f\right) \tag{1}$$

- **Running regression to estimate the CAPM beta:** Let $R_t^M$ be the month-$t$ return of the market portfolio and let $R^{GE}$ be the month-$t$ return of GE, and we run the following regression to estimate the CAPM beta:

$$R_t^{\mathrm{GE}} - r_f = \alpha + \beta \left(R_t^{\mathrm{M}} - r_f\right) + \epsilon_t \,. \tag{2}$$

  In terms of data structure, the above regression involves two time-series. In an Excel spreadsheet, they show up as two columns, one for $R^{GE}$ and the other for $R^M$. There is one more column, which is not explicitly given and usually you don't need to use it. It is the regression residual, $\epsilon_t$. So there are three random variables involved $R^{GM}$, $R^M$, and $\epsilon$, and two numbers to be estimated: $\alpha$ and $\beta$.

  This regression turns out to be a very important one in Finance, especially for understanding risk and returns in the equity market. So let's spend some time on it. First all of, this regression puts $R^M$ in a unique position. All other stocks and portfolios of stocks will show up on the left hand side as the dependent variable, while $R^M$ always sits on the right hand side as the independent variable. This is at the heart of the Bill Sharpe's insight: in the CAPM world, there is one risk that really matters ($R^M$), and everything else will be a reference to this unique portfolio. Second, the relation in the CAPM world is linear, and we are running a linear regression in agreement with that.

  Third, the regression coefficient $\beta$ is indeed the CAPM beta. This greatly simplifies the estimation process. Instead of doing $\mathrm{cov}(R^{GE}, R^M)/\mathrm{var}(R^M)$, we can simply run a regression to estimate $\beta$. Most importantly, the standard error of $\beta$ is given out for free as part of the regression output. Otherwise, you will be scratching your head to figure out how to calculate the standard error, because, so far, I've only taught you how to calculate the standard error of $\hat{\mu}$.

- **Running regression to estimate the CAPM alpha:** By running the regression, you also get an estimate of $\alpha$, which does not seem to be part of the CAPM. But in

fact, it is at the heart of the pricing relationship. Let's re-arrange the regression in Equation (2) to get:

$$\alpha = R_t^{GE} - r_f - \beta \left( R_t^M - r_f \right) - \epsilon_t \,. \tag{3}$$

Now let's take expectations on both sides to get:

$$\alpha = E \left( R_t^{GE} - r_f \right) - \beta \, E \left( R_t^M - r_f \right) - E(\epsilon_t) \,.$$

And what's $E(\epsilon_t)$? By construction, the regression residual, i.e., the time-series of $\epsilon_t$ always has zero mean. So we now have

$$\alpha = E \left( R_t^{GE} - r_f \right) - \beta \, E \left( R_t^M - r_f \right) .$$

So the CAPM pricing formula in Equation (1) is exactly as $\alpha = 0$. Isn't that neat?

Now the test of the CAPM is the same as testing whether or not $\alpha$ is zero. Remember how I kept emphasizing that any estimator will always inherit the noise/randomness in the data? Applying this lesson on the estimate for $\alpha$, we need to figure out its standard error. Luckily, this is given, again for free, as part of the regression output. Otherwise, what would you do?[4] Armed with the t-stat for $\alpha$, we can now easily test whether or not $\alpha$ is statistically significant.

Now you can appreciate why we keep including the clumsy $r_f$ in this regression. Its presence does not interfere too much with our estimate for $\beta$, because the volatility of $r_f$ is much small than that of stock returns. But by dragging $r_f$ along in this regression, we get $\alpha$! In other words, by running this regression, not only can we assess the risk exposure of a stock, but also its "abnormal" return relative to the CAPM benchmark.

- **CAPM dead or alive:** If we can create a lot of portfolios whose $\alpha$'s are significantly different from zero, then the CAPM, as least its version in Equation (1) will be in trouble. This is what happened to the CAPM in the 1990s, when academic researchers started to construct trading strategies that will give us positive alpha's. The most famous examples are the size and value portfolios in Fama and French (1992) and the momentum portfolios in Jegadeesh and Titman (1993). At the height of this research activities, some people even write papers with the title "The CAPM is Dead," which, in my opinion, is certainly an exaggeration because the CAPM is more than just Equation (1).

[4] In fact, $\alpha$ is like $\mu$ for the time-series of $R_t^{GE} - r_f - \beta \left( R_t^M - r_f \right)$. So a good approximation for the standard error of $\alpha$ is $\text{std}(\epsilon_t)/sqrt(N)$, where $\epsilon$ is the regression residual.

In fact, when Bill Sharpe was interviewed in 1998, he was asked exactly this question, "Some people proclaim that the CAPM is dead. What do you think?" He responded by saying that the insight of the CAPM is more than just the equation. Instead, the true insight about the CAPM is the risk that really matters:

> *"The fundamental idea remains that there's no reason to expect reward just for bearing risk. Otherwise, you'd make a lot of money in Las Vegas. If there's reward for risk, it's got to be special. There's got to be some economics behind it or else the world is a very crazy place. I don't think differently about those basic ideas at all." – Bill Sharpe, 1988, Dow Jones Asset Manager*

- **Running regression to get R-squared:** One last cool thing about running this regression is it gives us the R-squared of the regression. Going back to the regression in Equation 2, we see that it is a regression that links two random variables $R^{GE}$ and $R^M$. In an Excel spreadsheet, they show up as two columns. After the regression, we get a third random variable $\epsilon$, which is the residual of the regression. If you like, you can back it out as a column in the Excel spreadsheet by,

$$\epsilon_t = R_t^{\mathrm{GE}} - r_f - \alpha - \beta \left( R_t^{\mathrm{M}} - r_f \right) .$$

As you might know, by construction, the residual $\epsilon$ has mean zero and is uncorrelated with the independent variable: $\mathrm{cov}(R_t^M, \epsilon_t) = 0$.

Effectively, this regression decomposes $R^{GE}$ into two random components: the first one is associated with the market portfolio through the term $\beta\, R^M$, and the other is the regression residual $\epsilon$. This decomposition turns out to be very meaningful. In the CAPM language, the first term is the systematic component and the second is idiosyncratic component. And the total variance of $R^{GE}$ is the sum of these two components:

$$\mathrm{var}(R^{GE}) = \beta^2 \,\mathrm{var}(R^M) + \mathrm{var}(\epsilon) \,.$$

The R-squared gives us the ratio of how much of the variance in $R^{GE}$ comes from the systematic component:

$$\text{R-squared} = \frac{\beta^2 \,\mathrm{var}(R^M)}{\mathrm{var}(R^{GE})} \,,$$

and 1-R-squared gives us the ratio of how much of the variance in $R^{GE}$ comes from

the idiosyncratic component:

$$1 - \text{R-squared} = \frac{\text{var}(\epsilon)}{\text{var}(R^{GE})}\,.$$

R-squared are the most useful in telling us how important a risk factor is. For example, in the above regression, $R^M$ is used as a risk factor in explaining the variation in a stock or a stock portfolio like $R^{GE}$. A high R-squared indicates that the risk factor is very important in explaining the variation. For example, take any actively managed equity mutual fund returns and regress it on $R^M$, the typical R-squared is over 90%, with some funds very close to 98%. Later on, we will develop other risk factors such size, value, and momentum. Out of all the risk factors out there, $R^M$ remains to be the most important risk factor, second to none and by a wide margin. From this perspective, if there is risk factor that you should be paying attention to, it is the market portfolio $R^M$.

- **Why so much emphasis on this regression?** As you can see, I've spent quite some time focusing on just one regression. If before learning about this regression, the CAPM concepts seem vague and inaccessible, they should come alive to you by now. At least for myself, my understanding of the CAPM enhanced a great deal after having to teach students about this regression.

# Appendices

# A On Estimating the Expected Return

- **t-stat and Sharpe ratio:** You might notice that there is a connection between t-stat and Sharpe ratio. To make things more clear, let's assume that we are estimating the expected *excess* return. For this, we use a time-series of excess stock returns: $R_t - r_f$, where $r_f$ is the riskfree rate. For our current example, let's assume that the riskfree rate is a constant.

  Let $\text{avg}(R - r_f)$ denote the sample mean of this time-series, and $\text{std}(R)$ be the sample standard deviation. Then the t-stat of the estimator for the expected excess return is

  $$\text{t-stat} = \frac{\text{avg}(R - r_f)}{\text{std}(R)/\sqrt{N}} = \frac{\text{avg}(R - r_f)}{\text{std}(R)} \times \sqrt{N} = \text{Sharpe Ratio} \times \sqrt{N}\,.$$

  So if you go to an investment meeting, where t-stat's are usually not reported, you can use the reported Sharpe ratio to back out the t-stat. Moreover, if the Sharpe ratios of a wide range of products are reported, all with the same number of observations ($N$), then you know that the product with the highest Sharpe ratio also gives you the most significant (statistical speaking) expected excess return.

- **Estimating $\mu$ at higher frequencies:** Since the standard error of $\hat{\mu}$ depends on the number of observations (N), why don't we use the return data at a higher frequency? Well, it turns out that it doesn't really work. When it comes to the precision of $\hat{\mu}$, it is the length of the time-series that counts, not the number of observations. So chopping the time-series into finer intervals does not work. By contrast, when it comes to estimating the volatility of stock returns, this approach of chopping data into finer frequency does work and is widely used.

  Here is why for the first moment (i.e., mean), chopping does not work. Using N years of time-series of annual return data, we calculate the sample mean and sample standard deviation and use $\text{avg}(R)$ and $\text{std}(R)$ to denote them. So the t-stat at the annual frequency is

  $$\text{t-stat} = \frac{\text{avg}(R)}{\text{std}(R)/\sqrt{N}} = \frac{\text{avg}(R)}{\text{std}(R)} \times \sqrt{N}$$

  Now let's do things in a monthly frequency. The total number of observation increases by a factor of 12. As a pretty good approximation, moving from the annual to

monthly frequency, the sample mean becomes $\text{avg}(R)/12$ and the sample standard deviation becomes $\text{std}(R)/\sqrt{12}$. (Use log-returns and under the random walk model, this approximation becomes precise. We will re-visit this when we cover the Black-Scholes model.) So the t-stat at the monthly frequency is

$$\text{t-stat} = \frac{\text{avg}(R)/12}{\text{std}(R)/\sqrt{12}} \times \sqrt{N \times 12} = \frac{\text{avg}(R)}{\text{std}(R)} \times \sqrt{N}\,.$$

- **Other ways to estimate $\mu$:** Since $\mu$ plays such a uniquely important role in investments, you can rest assured that there are countless efforts in estimating/predicting this number. One student told me that in an interview, he was asked to design a derivative whose value would depend on $\mu$. My answer: it is not possible, because of the risk-neutral pricing (we will cover this later in the semester). There are also numerous surveys soliciting predictions about the markets from investors and economists. Well, when it comes to predicting the stock market, these survey data do not work very well (we will cover this topic later in the semester).

  Prof. Fama and French also had this interesting idea of estimating the expected return using dividend and earnings growth rates, which are much less noisy than stock returns. There is, of course, a pricing model that links stock prices to dividend or earnings, like the Gordon growth model you see in 15.415 or 15.402.

# B OLS Regression

## B.1 Introduction

In Finance, we run regressions left and right. That is, very often. Because of the availability of canned software routines, we have the luxury of not having to deal with the process that happens in the background. This is a pity, because the process itself is actually informative. So let me write this little note to add to the intuition behind an OLS regression. It is not as formal as what you will get from an Econometrics class, but adequate.

Code 1: My OLS Regression Code

```
function [out,R2]=Reg_OLS(Y,X)

A=[ones(length(Y),1) X];
b=inv(A'*A)*(A'*Y);
```

```
Eps=Y-A*b;
SE=sqrt(diag(inv(A'*A)*var(Eps)));

out=[b'; SE'; (b./SE)'];
R2=1-var(Eps)/var(Y);
% use this if need adjusted R2: adj_R2=R2-(1-R2)*size(X,2)/(size(X
   ,1)-size(X,2)-1);
```

Inserted above is my Matlab code for OLS regression. Throughout the semester, I use this little program to run regressions for all of the tables and figures in the lecture slides and notes. The inputs are the regressand $Y$ and the regressor $X$. If these names are confusing (they are to me), then let's call them dependent variable $Y$ and independent variable $X$. Or even simpler, left-hand side and right-hand side variables. The $Y$ variable is always one-dimensional: $N \times 1$, where $N$ is the number of total observations. The $X$ variable is of dimension $N \times k$, where $k$ is the number of independent variables (or explanatory variables).

For the CAPM regression in this class:

$$R_t^i - r_f = \alpha + \beta \left( R_t^M - r_f \right) + \epsilon_t^i \,,$$

I can feed the program with $R_t^i - r_f$ as $Y_t$ and $R_t^M - r_f$ as $X_t$ and the program will give me the regression output: estimate, standard error, t-stat, and R-squared. As you can see, the matrix operation makes the formula quite simple, especially when we have more than one explanatory variable. For example, as we expand the CAPM setting to the Fama-French three-factor model, we will have three explanatory variables (market, size, and value).

## B.2 A Concrete Example

One limitation of the matrix operation is that it is not very transparent. So let's work with an example with just one explanatory variable. To sharpen our focus, let me further assume that both $Y$ and $X$ have zero mean. This way, we don't have to deal with $\alpha$, and can focus only on $\beta$:

$$Y_t = \beta \, X_t + \epsilon_t \,,$$

- **The Regression Coefficient $\beta$:** In running a regression, the mathematical program we are trying to solve is in fact a linear prediction problem. The goal is to minimize the prediction error:

$$\min_{\beta} \sum_t \left( Y_t - \beta \, X_t \right)^2$$

If you are good at solving optimization problems, it is pretty easy to see that the first order condition of the above optimization problem gives us the solution:

$$\hat{\beta} = \frac{\sum_t Y_t X_t}{\sum_t X_t^2} .$$

Recall that $X$ and $Y$ are zero mean (for simplicity). So

$$\frac{1}{N}\sum_t Y_t X_t = \text{cov}\,(Y_t, X_t) \; ; \quad \frac{1}{N}\sum_t X_t^2 = \text{var}\,(X_t) \; .$$

Now you can see why we can recover the CAPM $\beta$ by running a regression.

- **The Standard Error:** Note that I've put a hat on $\beta$ to emphasize that this is an estimator, inheriting the noise from the data. So it is a random variable itself (just like $\hat{\mu}$). Now let's calculate its standard error, which is the square-root of

$$\text{var}\left(\hat{\beta}\right) = \text{var}\left(\frac{\sum_t Y_t X_t}{\sum_t X_t^2}\right) = \frac{\text{var}\,(\epsilon_t)}{\sum_t X_t^2} ,$$

where I've abused the notation a bit. To be more precise, I should use $\text{var}\left(\hat{\beta} \,|\, X\right)$. In other words, I am doing the calculation conditioning on $X$ and taking advantage of the result that the residual $\epsilon_t$ is by construction independent of $X_t$.

Using $\sigma_\epsilon$ for the variance of the residual and $\sigma_X$ for the variance of $X$ (and remember that both $Y$ and $X$ have zero mean), we can further simplify the standard error to

$$\text{var}\left(\hat{\beta}\right) = \frac{\sigma_\epsilon^2}{N\,\sigma_X^2} ; \quad s.e.(\hat{\beta}) = \frac{\sigma_\epsilon}{\sqrt{N}\,\sigma_X} .$$

Going back to the intuition we've gained by working with $\hat{\mu}$, we know that the longer the time-series (larger $N$), the more precise the estimator. Here we have the same result: with more observations (larger $N$), the standard error for $\hat{\beta}$ is smaller. Also interesting is the fact that the residual variance $\sigma_\epsilon^2$ has a direct impact on the precision of $\hat{\beta}$. Recall that the noisier the stock market (higher $\sigma_R$), the less precise the estimator $\hat{\mu}$. Here it is the ratio of $\sigma_\epsilon/\sigma_X$ that matters: for a given level of $\sigma_X$, the noisier the residual (the unexplained component), the less precise the regression coefficient $\beta$.

- **The More General Case:** In making the example concrete, we've assumed that $X$ and $Y$ are zero mean random variables. Taking out this assumption, we run the more

general regression of

$$Y_t = \alpha + \beta\, X_t + \epsilon_t\,.$$

The two important results we've obtained so far remain true:

$$\hat{\beta} = \frac{\mathrm{cov}(Y_t, X_t)}{\mathrm{var}(X_t)}\,; \quad s.e.(\hat{\beta}) = \frac{\sigma_\epsilon}{\sqrt{N}\,\sigma_X}\,.$$

Moreover, you will find the quite intuitive result of

$$\hat{\alpha} = E(Y_t) - \hat{\beta}E(X_t)\,; \quad s.e.(\hat{\alpha}) = \sqrt{\frac{\sigma_\epsilon^2}{N} + \Big(s.e.(\hat{\beta})\, E(X_t)\Big)^2}\,.$$

From this solution, you can see that $\hat{\alpha}$ is very similar to the mean estimator $\hat{\mu}$. If the explanatory variable $X$ is zero mean, then the estimation for $\alpha$ will not involve $\beta$. In this case, $\hat{\alpha}$ is indeed the mean estimator for the residual $\epsilon_t$. Applying this to the CAPM setting, where $\sigma_\epsilon$ is the volatility of the idiosyncratic risk taken by a portfolio manager, you can see how the precision of his $\alpha$ is linked to his level of idiosyncratic risk. Having to estimate his $\beta$ adds a bit noise to the precision of $\alpha$, but it is the amount of the idiosyncratic risk that is the main driver for $s.e.(\hat{\alpha})$. So if a manager achieves his $\alpha$ through exposing his portfolio to high idiosyncratic risk, his signal to noise ratio will be low and the precision of his $\alpha$ will also be low.

- **R-squared and its relation to t-stat:** The R-squared of a regression provides additional information. Going back to the CAPM regression, we notice that two stocks with the same $\beta$ could have very different R-squared's. You can dial up the idiosyncratic risk to decrease the R-squared while keeping the same $\beta$. But it turns out that there is a one-to-one relation between R-squared and t-stat:

$$\text{R-squared} = \frac{\beta^2\,\sigma_X^2}{\sigma_Y^2}$$

and

$$(\text{t-stat})^2 = \left(\frac{\hat{\beta}}{s.e.(\beta)}\right)^2 = \frac{N\,\beta^2\,\sigma_X^2}{\sigma_\epsilon^2}$$

Comparing the two and using the fact that

$$R^2 = \frac{\beta^2\sigma_X^2}{\sigma_Y^2}\,; \quad 1 - R^2 = \frac{\sigma_\epsilon^2}{\sigma_Y^2}$$

we have

$$(\text{t-stat})^2 = \frac{N\,R^2\,\sigma_Y^2}{\sigma_\epsilon^2} = \frac{N\,R^2}{1-R^2}\,,$$

where, to make the math look pretty, I've used $R^2$ for R-squared. So if I give you two stocks with the same t-stat for $\beta$ (using the same number of observations $N$), their R-squared must be the same.

- **One final note:** This note is not really necessary. It addresses the issue regarding unbiasness in small sample and consistency in large sample. I am writing it simply for those who notice that sometimes we scale the number of observations by $N$, sometime $N-1$, and sometimes $N-2$.

  Suppose you are given $N$ observations of a random variable $Z$ and you are asked to estimate its variance. You must have been taught in one of your earlier statistics classes that the unbiased estimator is

$$\frac{1}{N-1}\sum_{t=1}^{N}\left(Z_t - \bar{Z}\right)^2$$

  Notice the term $N-1$ (not the usual $N$) is used as the scaling factor. Although we have a total of $N$ observations, the degree of freedom is only $N-1$ because by having to estimate the mean of $Z$ (i.e., $\bar{Z}$), we use up one degree of freedom. You can go through the very tedious algebra (and I am sure you were asked to do so in your Statistics class) to convince yourself that this is indeed true.

  I have to confess that I am not super crazy about making this adjustment from $N$ to $N-1$ to get an unbiased estimator. For large $N$, the difference between $N-1$ and $N$ is negligible. If we take $N$ to infinity, then adding or subtracting a finite number from it will not matter. So I very much prefer to use

$$\frac{1}{N}\sum_{t=1}^{N}\left(Z_t - \bar{Z}\right)^2$$

  as an estimator. In the language of Econometrics, this is a consistent estimator: it converges to the true value when the sample size $N$ grows to infinity.

  By the way, to do this adjustment for OLS regression, the degree of freedom is $N-2$ for the case of one explanatory variable. We are losing two degrees of freedom because in order to estimate the variance of the residual ($\sigma_\epsilon$) we need to first estimate the

intercept and the slope coefficient. So the standard error for $\hat{\beta}$ would be

$$s.e.\,(\beta) = \frac{\sigma_\epsilon^2}{N\sigma_X^2} = \frac{\sum_t \epsilon_t^2/(N-2)}{N\sigma_X^2}$$

On the other hand, this bias adjustment is not done in calculating the R-squared. Consequently, the relationship between t-stat and R-squared would be

$$(\text{t-stat})^2 = \frac{(N-2)\,R^2}{1-R^2}\,.$$

# C Exercises

I will add more if possible. I also welcome suggestions from students.

1. In a paper written by Prof. Jiang Wang and his co-authors, it was reported that for the Chinese market from July 1997 to December 2013, the average excess return of the market portfolio is 0.60% per month with a t-stat of 0.97. What is the monthly stock market volatility in China during this sample period? Compare it with the US market volatility.

2. Can two stocks have the same $\beta$ but different R-squared? If so, construct an example for me? In which way are these two stocks similar? In which ways are they different?

3. Can two stocks have the same t-stat of $\beta$ but different R-squared?

4. Suppose that there are many different companies whose stocks have the same beta, say 1. Can you form a portfolio to diversify the risk to get a lower beta?

5. Suppose you have a put option on the S&P 500 index. Do you think it has a non-zero beta? If so, it is positive or negative?

# D Matlab Code

Code 2: Plot the Distribution of $R_t$ and $\hat{\mu}$

```
Mu=0.12; Sigma=0.20; N=88;

SE=Sigma/sqrt(88);
```

```
R_Grid=-0.40:0.0005:0.50;
PDF_R=normpdf(R_Grid,Mu,Sigma);
PDF_MuHat=normpdf(R_Grid,Mu,SE);

figure(1);clf;
plot(R_Grid*100,PDF_R,'b-',R_Grid*100,PDF_MuHat,'r-');
text(0.21*100,16,'\bf pdf of $$\hat{\mu}$$','Interpreter','Latex','FontSize'
    ,13)
ylabel('\bf Probability Density');
xlabel('\bf Return (%)');
text(-0.2*100,3.0,'\bf pdf of $$R_t$$','Interpreter','Latex','FontSize',13);

text(-0.33*100,16,'\bf $$\hat\mu = \frac{1}{N} \sum_{t=1}^N R_t$$','
    Interpreter','Latex','FontSize',13);
title('\bf The 95-Percent Confidence Interval');
PDF_R=normpdf(R_Grid,Mu,Sigma);
PDF_MuHat=normpdf(R_Grid,Mu,SE);
Conf95_Left=Mu-1.96*SE;
Conf95_Right=Mu+1.96*SE;
Left=(Conf95_Left:-0.001:Mu-4*SE);
Right=(Conf95_Right:0.001:Mu+4*SE);
pdf_Left=normpdf(Left,Mu,SE);
pdf_Right=normpdf(Right,Mu,SE);
hold on;
fill([Left fliplr(Left)]*100,[zeros(1,length(pdf_Left)) fliplr(pdf_Left)],'y
    ');
fill([Right fliplr(Right)]*100,[zeros(1,length(pdf_Right)) fliplr(pdf_Right)
    ],'y');

hold off;
```

Jun Pan | MIT Sloan School of Management | 253-3083
junpan@mit.edu | 15.433-15.4331 Financial Markets, Fall 2015 | E62-624

## Classes 4 & 5: Equity in the Cross Section, Part 1

This Version: September 26, 2016

When I presented the timeline of Modern Finance in class, one of the students asked why the blue events (work developed by academics) stopped after the 1970s. Of course, we academics kept writing papers, at an even faster rate. But the 1970s were a great time to do Finance in the academic world. Theoretical papers that are foundation building and trail brazing happened in that era. Since the 1990s, most of the exciting work in Finance happened in the empirical area.

What we are going to cover in this class represents the most influential work in Finance since 1990s. And the intellectual leader is Prof. Eugene Fama, who was awarded a Nobel prize in 2013. The ideas behind the research papers helped inspire and create this fast growing area called quant investing in the late 1990s and early 2000s. More recently, with the growing popularity of factor investing, the mutual fund and ETF world is also incorporating these ideas.

As a PhD student at Stanford GSB in the late 1990s, I didn't know much about the Fama-French factors. I was into my own research at that time. Fortunately, I had to teach 15.433 at MIT Sloan. So it was through having to teach the MBA students at Sloan that I got to learn and admire the work of Prof. Fama and his co-authors.

# 1 Quant Investing

Both quant investors and stock pickers are interested in generating alpha, but they differ in their approach. To argue which approach, stock picker or quant investing, is better is meaningless, but to find out which one suits you better is extremely important. To quote a recent column by John Authers from the *Financial Times*, "If we do want to try to do better than passive then there are two logical ways to do it. Either we can adopt a tightly disciplined approach designed to exploit persistent market anomalies or factors; or we can focus tightly on a sector or industry and make concentrated bets with high conviction." So on the one end of the spectrum of this alpha generating business is an investor like Warren Buffett, who makes concentrated bets with high conviction; and on the other end are many

long/short equity quant funds using quant signals to exploit persistent patterns in the cross-section of stocks. The Warren Buffett path has clearly been admired and traveled by many but there are limited number of success stories. After all, how can you replicate a person's mind? The quant approach, on the other hand, has generated relatively more success stories. This approach serves more as a tool and it is much easier to replicate. Not surprisingly, it ended up as an over-crowded field. In terms of skills, quant investing really does not require a lot of quantitative skills in the traditional sense. It requires a curious and creative mind, love and respect for the data, and some basic programming skills such as regressions and data cleaning, merging, and sorting.

The first observation of quant investing is that even within the US equity markets, there are thousands of stocks to choose from. For a stock picker, it would be a daunting task to cover them all. With the availability of computers and data, it seems obvious that they should develop a systematic approach to search through the data for alpha. This, of course, assumes that the patterns found in the data persist in the near future. This is where quantitative signals come in. The key insight is that such quant signals are useful in separating one group of stocks (high alpha) from another (zero or negative alpha). Potentially, there are two interpretations or reasons as to why such signals might work. First, they help us exploit the mis-pricing in the markets. Second, they represent differences in exposure to certain risk factors (that are unrelated to the market portfolio). Subscribing to the first interpretation, you believe that your alpha comes from market inefficiency. The second line of reasoning leads you to believe that your alpha comes from exposures to certain systematic risk (that is unrelated to the market portfolio). In this case, the alpha's are simply beta's in disguise.

One signature approach of quant investing is forming portfolios. This arises from the desire to be exposed only to the risk (or anomaly) one is interested in. The portfolio approach helps diversify away unwanted idiosyncratic risk. Another signature approach of quant investing in the hedge fund world is the long/short strategy. Again, this arises from the desire to have a razor sharp focus on the target risk factor. The long/short strategy helps take out the unwanted systematic risk (e.g., the market risk). The best place to learn about quant investing is to read carefully the tables in Fama and French (1992, 1993). Afterwards, go to Prof. French's website and play with the data.

The most creative part of quant investing is to come up with signals that could generate alpha, especially those signals that help us identify market inefficiency in the cross-section. Unfortunately, most of the signals used by quant funds have their origin in academic papers and, in my opinion, are not that creative. It either indicates that markets are not that inefficient, or quant funds are not that creative.

The need for more innovation in this area certainly shows up during the recent quant meltdown in August 2007. What we learned from this event is that the quant investing space is very crowded, populated by funds with very similar ideas. The initial success of quant investing in the 1990s attracted many investors, and the quant investing world enjoyed a great rise in the first half of 2000s.[1] It turns out that many quant funds are trading on very similar signals. Prof. Daniel, who was at Goldman during the quant meltdown, wrote an interesting and informative set of slides. The initial trigger was in the sub-prime mortgage market, and it spilled over to investment-grade credit markets shortly thereafter. As multi-strategy hedge funds experiences losses in their illiquid mortgage and credit positions, they liquidated their more liquid assets in the quant investing side to raise cash. As this unwinding took place, many quant investing funds rushed to the door, triggering a 20-sigma move in the quant investing space. Previously unrelated stocks suddenly started to move together during the unwind. If you were not in the quant space, you probably would not have noticed the 20-sigma move. But if you are in the quant space, then most likely your portfolio experienced a 10 to 20 sigma drop over one week.

In recent years, the basic ideas in quant investing have found their popularity in the world of mutual funds and ETFs. While the sales pitch in the quant hedge fund world is all about Alpha, now the emphasis is on Beta: smart beta and factor investing. In any case, if you are interested in a career in this area, what we are going to cover in the next few classes is going to be very useful. Coming straight from the original research papers, it is also the gold standard.

# 2 Forming Portfolios using Quantitative Signals

- **Popular quant signals:** Quant investing uses stock characteristics as signals. Most quant investors believe that their signals help capture the fundamentals that drive alpha. Here is a list of widely adopted quant categories and strategies:
  - Size: The market capitalization = stock price $\times$ number of shares outstanding.
  - Valuation: How is the company priced relative to fundamental accounting measure? For this, we have the widely used book-to-market ratio:

$$\texttt{BtM} = \frac{\text{book value of equity}}{\text{market value of equity}}.$$

[1] I often infer the popularity of a field from the number of Finance professors (whom I personally know) it managed to attract to switch jobs. In mid-2000s, I observed quite a few.

- Momentum: How has the market responded to the company's changing fortunes? Sample Metric: `price momentum`.
- Profitability: What are the company's profit margins? How efficient are its operations? Sample metric: `earnings-to-sales ratio` and `OP` (operating profitability) in the five-factor model of Fama and French.
- Earnings Quality: Were earnings derived from sustainable sources? Sample metric: `the accruals-to-total-assets`.
- Analysts Sentiment: Are analysts upgrading or downgrading their view of this company? Sample metric: `earnings forecast revisions`.
- Management Impact: How is the company's management employing its capital? Sample metric: `change in shares outstanding` or the `Investment` variable (growth in firm assets) in the five-factor model of Fama and French.

In coming up with the above list, I mostly used the information from Prof. Daniel's slides. In addition, I also listed the two new Fama-French variables, Profitability and Investment, from their recent five-factor model. We will cover size, value, and momentum in detail. For most of the other signals, Googling will lead you to the key research articles behind these strategies.

These signals differ in various ways. Some are momentum signals (e.g., earnings forecast revisions), indicating a slow reaction to information. Some are contrarian signals (e.g., valuation), indicating a reversal in price pattern due to over-reactions in the past. Some are over a long horizon. For example, studies on change in shares outstanding (due to seasoned equity offerings or share repurchase announcements) focus on returns with holding periods of 3 or more years. Some are over a horizon of a few months (e.g., momentum).

- **Sorting stocks into portfolios:** The concept of sorting is pretty straightforward. Of course, there are many details one needs to pay attention to. The best resources are Fama and French (1992), which by now is the gold standard in this area. Prof. French's website also provides a great deal of information. It should be mentioned that Prof. French offers a tremendous service to our profession by making the data available on his website. If I didn't have access to the materials posted on his website, I would have to construct a lot of the tables and plots in this class from scratch.

  In this class, we will first look at univariate sorts (by size or book-to-market) into deciles, and then move on to double sorts (by size and book-to-market) into 5x5. One

important convention is that the breakpoints of these sorts are first established by using NYSE stocks only. The main reason is that the stock population in NYSE is more representative.

It is also important to emphasize that sorting is done dynamically. Stock characteristics fluctuate over time. So we need to periodically update this information and re-sort stocks by their new characteristics so that the sorted portfolios contain stocks of the right characteristics. The frequency of sorting depends on the variability of the signals. For example, Fama and French sort their size portfolios once a year, using the market value in June of year t for portfolio returns from July of year t to June of t+1. For the momentum portfolios, however, the signal is the stock's past returns, which are more variable, and the sorting is done at a monthly frequency. More generally, the variability of a signal also affects the portfolio turnover. For a signal such as market cap, the portfolio turnover is low because market cap is relatively stable. By contrast, for a momentum signal such as past stock returns, the portfolio turnover could be quite high. All of these considerations could factor into the execution costs of a strategy.

- **Size and BtM sorted portfolios:** The size-sorted deciles are useful in our understanding of the overall size distribution of stocks listed on the three US exchanges. Using the 2015 number, we see that the average market cap is a mere $116 millions for stocks in decile 1, which contains 1362 stocks. By contrast, decile 10 has only 173 stocks. Given that the breakpoints are determined by NYSE stocks, this implies that most of the AMEX and Nasdaq stocks fall into the smaller deciles. For stocks in decile 10, the average market cap is close to $84 billions. Of course, this is still no comparison to those mega-large stocks such as Google ($427B), Apple ($661B), or Amazon ($244).

  The book-to-market sorted deciles give us a sense of how much the equity value of a firm differ from its book value. For some stocks, equity investors value the stocks to the extent that they are willing to pay much more than the existing book value of its equity. As a result, the market value of the equity takes into account the firm's future growth component, which is not reflected in the firm's current book value. Such growth stocks are of low book-to-market ratio and show up in the lower deciles. For example, using the 2015 number, the average book-to-market ratio of stocks in decile 1 is 0.095: for each dollar in market value, the book value is only 0.095. Or, for each dollar in the book value, the market is willing to pay 1/0.095=10.5 dollars. You can imagine that Google was once a growth stock. Back in 2006, Google had a book-to-market ratio of 0.04 and its market cap was $107B. Right now, its price-to-book is 3.84 according to Yahoo Finance. So its book-to-market ratio has gone up quite a bit in the past 10

years to it current level of 0.26.

At the other end of the spectrum are stocks with very high book-to-market ratio. These stocks, usually referred to as value stocks, have a depressed market value. Using the 2015 number, we see that stocks in our decile 10 have an average book-to-market ratio of 1.339.[2] Basically, investors are not willing to pay the full book value for the stock. For example, back in 2006, the book-to-market ratio of GM was 1.28. For each dollar in book value, investors are only willing to pay only 1/1.28=0.78 dollar in the stock market. On the morning of June 1, 2009, GM filed for bankruptcy protection. In general, firms have a high book-to-market ratio prior to filing for bankruptcy, but this does not mean that high book-to-market firms are bankruptcy firms.

At this point, it is worthwhile to emphasize again that sorting is done dynamically. For example, back in 2006, GM, with its book-to-market value of 1.28, showed up in the book-to-market decide 10. After its filing for bankruptcy protection, GM dropped out of the sample. As of today (September 14, 2015), according to Yahoo Finance, GM has a price-to-book ratio of 1.37, indicating a book-to-market ratio of 1/1.37=0.7299. So now GM shows up in decile 7 or 8.

# 3 Testing the CAPM using Fama-French 25 Portfolios

Let's start with the regression:

$$R_t^i - r_f = \alpha_i + \beta_i \left(R_t^M - r_f\right) + \epsilon_t^i \,, \tag{1}$$

where $R_t^i$ is the month-$t$ return of a portfolio $i$. Recall that testing the CAPM pricing equation is equivalent to testing whether or not $\alpha_i$ is significantly different from zero. If we can find many portfolios with large $\alpha$'s, then the CAPM will be in trouble. Indeed, this is at the heart of what we are going to do.

- **Use the CAPM beta:** We use the famous Fama-French 25 portfolios to test the CAPM. For each portfolio $i$, we run the regression in Equation (1) to obtain its $\beta_i$. After obtaining an estimator for the market risk premium $\lambda^M$, we calculate the risk premium for portfolio $i$ according to the CAPM: $\beta_i \lambda^M$. We call this number the risk premium predicted by the CAPM. At the same time, we use the realized returns of portfolio $i$ to estimate the risk premium directly. We call this number the risk premium

[2]The reported average BtM is value weighted. That is, within each decile, we value-weight each stock's book-to-market ratio by its size to calculate the average BtM for the decile.

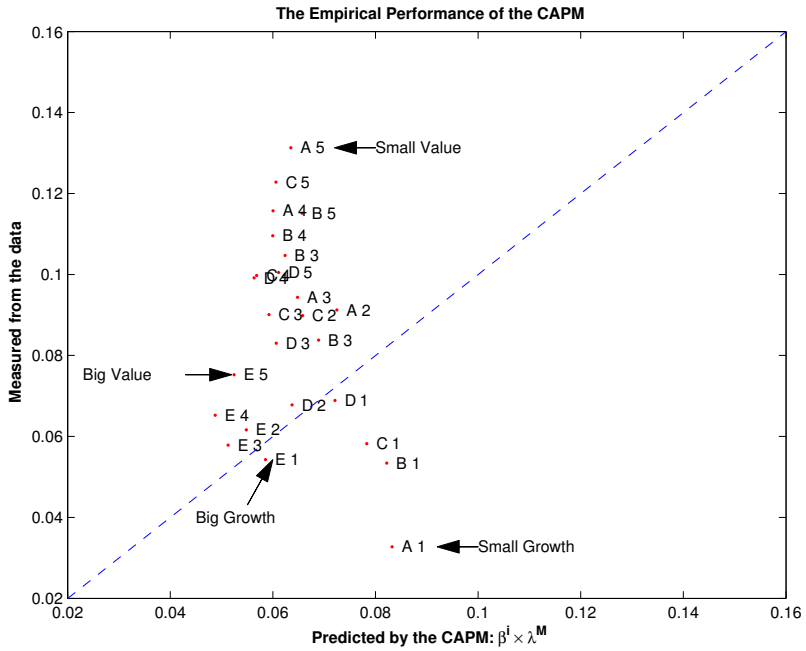


Figure 1: The empirical performance of the CAPM, using the Fama-French 25 portfolios. For each portfolio $i$, its risk premium measured from the data is plotted against that predicted by the CAPM.

measured from the data. We now have 25 pairs of numbers, each pair corresponds to one of the Fama-French 25 portfolios.

Figure 1 plots the 25 pairs of numbers: the risk premium measured from the data (y-axis) and the risk premium according to the CAPM (x-axis). I understand that both pairs of numbers are noisy because they are estimated from the data. But let's use them for now, and we will come back to a proper test later. Also, although the estimations and regressions are all done using monthly returns, I annualized the risk premium (by multiplying the monthly risk premiums by 12) for ease of communication.

Now let's come back to Figure 1. If the CAPM works well, then these 25 dots should line up pretty nicely along the 45-degree line: data and model in agreement. In practice, however, most of these dots are clustered together along the x-axis dimension and spread out along the y-axis dimension. Recall that plotted along the x-axis is model-implied risk premium: $\beta_i \lambda^M$. So effectively, the "clustering" implies that most of the 25 portfolios have very similar $\beta$. Moreover, given that the market risk premium $\lambda^M$ is close to 6% per year, this implies that most of the 25 portfolios have a $\beta$ that is very close to one. On the other hand, the wide variation along the y-axis dimension indicates that those portfolios in fact perform very differently in reality: some perform

well with a high risk premium, while some perform poorly with a low risk premium.

Overall, Figure 1 is not good news for the CAPM. Instead of the predicted relation between risk premium and beta, we find a wide range of risk premiums for portfolios that are very similar in beta. At this point, you might ask how much we should trust those 25 pairs of numbers, which are estimated with noise. So let's do our test more properly. Recall that the key to the test is Alpha. We can estimate the alpha's from the plot by measuring the vertical distance between each dot and 45-degree line. From the plot, we see that some portfolios (mostly value stocks) have positive alpha's, while other portfolios (mostly growth stocks) have negative alpha's.

- **Use the CAPM alpha:** Recall that the alpha's can be easily obtained from the regression in Equation (1). Table 1 reports, for the 25 portfolios, the CAPM $\alpha$, $\beta$, and the adjusted R-squared from the 25 regressions. For those portfolios with statistically significant $\alpha$'s, I print the number in bold. The t-stat's of the $\alpha$'s are also reported in the table. According to the CAPM, all of the $\alpha$'s should be indistinguishable from zero. But we have quite a few portfolios with statistically significant $\alpha$'s. Moreover, there is a pattern to it. For example, for small stocks in group A, moving along the book-to-market dimension, the portfolio $\alpha$'s turned from negative (and statistical significant) to positive. The same pattern repeated for all size groups. For large stocks in group E, none of the $\alpha$'s are statistically significant, but you can see the magnitude of $\alpha$'s increasing as we move the book-to-market from low to high.

  To jointly test the statistical significance of those 25 $\alpha$'s, we can use the GRS test, named after Gibbons, Ross, and Shanken (1989). It is actually a very cool test. It maps the joint $\alpha$ test to how inefficient the market portfolio $R^M$ is. If you recall, the CAPM tells us that the market portfolio is the tangent portfolio sitting at the mean-variance frontier with the highest Sharpe ratio. By being able to construct portfolios with positive alpha's, the story breaks down: the market portfolio is no longer the mean-variance efficient portfolio.

- **The importance of the CAPM:** Before closing this section, I would like to emphasize one more time that this test result does not hurt the importance of the CAPM model in Finance. In fact, without the model, we will not even know where and how to start the test. Moreover, the later development, including the Fama-French models we will see, always includes the market portfolio in the test. Indeed, the CAPM model serves as the foundation for all models to come. I have yet to see one model without the market portfolio in it. Finally, the main insight of the CAPM remains: there are

Table 1: The Fama-French 25 Portfolios in the CAPM and the Fama-French Three-Factor Model. All $\alpha$'s are reported in annualized terms (x12). Statistically significant $\alpha$'s are reported in bold. Monthly data from January 1962 to July 2015.

| | CAPM | | | | The FF Three Factor Model | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Portfolio | $\alpha$ (%) | t-stat | $\beta$ | R2 (%) | $\alpha$ (%) | t-stat | $\beta$ | $s$ | $h$ | R2 (%) |
| A1 | **-5.05** | -2.19 | 1.41 | 62.81 | **-5.32** | -4.69 | 1.06 | 1.38 | -0.29 | 91.25 |
| A2 | 1.88 | 0.95 | 1.23 | 63.50 | -0.10 | -0.13 | 0.96 | 1.30 | 0.04 | 94.20 |
| A3 | 2.95 | 1.80 | 1.10 | 66.77 | -0.09 | -0.15 | 0.92 | 1.10 | 0.28 | 95.20 |
| A4 | **5.57** | 3.46 | 1.02 | 64.34 | **1.65** | 2.57 | 0.89 | 1.03 | 0.46 | 94.48 |
| A5 | **6.78** | 3.82 | 1.08 | 62.38 | **1.49** | 2.21 | 0.98 | 1.09 | 0.70 | 94.71 |
| B1 | -2.88 | -1.68 | 1.39 | 74.93 | **-2.09** | -2.73 | 1.11 | 0.99 | -0.39 | 95.10 |
| B3 | 1.49 | 1.08 | 1.17 | 76.33 | -0.42 | -0.62 | 1.01 | 0.87 | 0.13 | 94.35 |
| B3 | **4.23** | 3.27 | 1.06 | 75.07 | 1.07 | 1.63 | 0.97 | 0.77 | 0.39 | 93.74 |
| B4 | **4.96** | 3.78 | 1.02 | 73.07 | 0.89 | 1.43 | 0.97 | 0.73 | 0.56 | 94.15 |
| B5 | **4.94** | 3.06 | 1.11 | 68.12 | -0.66 | -1.00 | 1.08 | 0.87 | 0.81 | 94.77 |
| C1 | -2.01 | -1.41 | 1.33 | 79.58 | -0.60 | -0.84 | 1.09 | 0.73 | -0.44 | 95.03 |
| C2 | **2.40** | 2.23 | 1.12 | 82.83 | 0.67 | 0.85 | 1.04 | 0.53 | 0.18 | 91.10 |
| C3 | **3.08** | 2.83 | 1.00 | 79.31 | 0.08 | 0.10 | 0.99 | 0.44 | 0.44 | 89.73 |
| C4 | **4.29** | 3.68 | 0.96 | 75.49 | 0.38 | 0.50 | 1.00 | 0.40 | 0.62 | 90.18 |
| C5 | **6.22** | 4.31 | 1.03 | 69.61 | 1.23 | 1.44 | 1.06 | 0.55 | 0.77 | 89.58 |
| D1 | -0.32 | -0.30 | 1.22 | 85.24 | **1.46** | 2.05 | 1.06 | 0.38 | -0.42 | 93.73 |
| D2 | 0.40 | 0.45 | 1.08 | 86.89 | -1.03 | -1.25 | 1.08 | 0.22 | 0.21 | 89.15 |
| D3 | **2.24** | 2.21 | 1.03 | 82.26 | -0.44 | -0.52 | 1.08 | 0.18 | 0.45 | 88.26 |
| D4 | **4.28** | 3.96 | 0.96 | 77.91 | 0.85 | 1.09 | 1.02 | 0.22 | 0.57 | 88.79 |
| D5 | **3.94** | 2.81 | 1.04 | 71.14 | -0.84 | -0.89 | 1.14 | 0.25 | 0.81 | 87.45 |
| E1 | -0.43 | -0.56 | 0.99 | 88.52 | **1.88** | 3.44 | 0.98 | -0.24 | -0.36 | 94.19 |
| E2 | 0.68 | 0.91 | 0.93 | 87.53 | 0.47 | 0.71 | 0.99 | -0.22 | 0.09 | 90.24 |
| E3 | 0.66 | 0.70 | 0.87 | 79.61 | -0.65 | -0.83 | 0.97 | -0.23 | 0.30 | 86.20 |
| E4 | 1.65 | 1.50 | 0.83 | 71.88 | **-1.38** | -2.03 | 0.98 | -0.20 | 0.60 | 89.41 |
| E5 | 2.28 | 1.57 | 0.89 | 62.79 | -1.76 | -1.65 | 1.05 | -0.08 | 0.76 | 80.48 |

undiversifiable risks in the market, and you get rewarded for bearing this kind of risk. There is no reward for holding diversifiable risk. In fact, it is this insight that prompts people to locate new risk factors. The one-factor structure of the CAPM might not work very well with the data, and the new multi-factor models might work better. But all builds on this insight of locating systematic risk factors.

# 4 The Fama-French Three Factor Model

Our test of the CAPM informs us on how the CAPM failed to price the Fama-French 25 portfolios: value stocks outperform growth stocks, and small stocks outperform big stocks. In the one-factor model of CAPM, the only risk factor is the market portfolio and the only measure of risk is beta. There is no additional role for size or value in the model. So the logical next step is to build a model that incorporates these two factors. This is what Fama and French did in their 1993 paper by introducing the SMB and HML factors.

- **The Fama and French factors:** In order to construct the factors, Fama and French use a coarser double sort. Along the size dimension, stocks are sort into two groups: small or big. Along the value dimension, stocks are sort into three groups with 30% in value, 40% in neutral, and 30% in growth. Because these portfolios are to be used to construct factors, one would like to have them as diversified as possible. A coarser sort would allow each bin to have more stocks and therefore improve diversification. Using the 6 (2x3) portfolios, the SMB and HML factors are constructed as
  - SMB (Small Minus Big):

$$R^{\text{SMB}} = R^{\text{small}} - R^{\text{big}},$$

    where $R^{\text{small}}$=1/3 (small value + small neutral + small growth) and $R^{\text{big}} = 1/3$ (big value + big neutral + big growth)
  - HML (High Minus Low):

$$R^{\text{HML}} = R^{\text{value}} - R^{\text{growth}},$$

    where $R^{\text{value}}$=1/2 (small value + big value) and $R^{\text{growth}}$=1/2 (small growth + big growth).

  As you can see, the factors are constructed by a long/short strategy. For example, the HML factor involves buying value stocks and selling growth stocks. The motivation

behind such a factor is to help investors focus on the targeted risk factor, which is the difference between values and growth stocks. Any unwanted risks are taken out from the factor: the portfolio approach diversifies away the idiosyncratic risk in individual stocks and the long/short strategy hedges out the exposure to the market risk.

As you might notice, SMB and HML are not totally orthogonal to the market risk. The beta's of small stocks are usually higher than those of big stocks. As a result, SMB has a slightly positive beta (around 0.20). The beta's of growth stocks are usually higher than those of value stocks. As a result, HML has a slightly negative beta (around $-0.2$). A purist would not like this. Nevertheless, by choosing to form their factors using such a simple long/short strategy, Fama and French seem to value simplicity and intuitiveness over perfection. I would have done the same thing given the cost and benefit.

- **The three-factor regression:** Now we are ready to run the following regression for our 25 portfolios:

$$R_t^i - r_f = \alpha_i + \beta_i \left(R_t^M - r_f\right) + s_i \, R_t^{\text{SMB}} + h_i \, R_t^{\text{HML}} + \epsilon_t^i \tag{2}$$

Notice that we've put the two new factors in the regression and label the corresponding slope coefficients to be $s$ and $h$. If you like, you can think of them as the "beta" on SMB and HML.

Since the new factors are slightly correlated with the existing factor, $R^M - r_f$, the $\beta$ in the current regression is no longer the CAPM beta. In fact, using Table 1, we can see that the $\beta$ from this new regression are slightly different from the CAPM $\beta$. The benefit of using SMB and HML is that they are very simple to construct and also very intuitive. As we will work in a regression framework for the three factor model, having slightly correlated factors is not a problem at all. Just be careful with the interpretation of the new $\beta$.

Table 1 also reports the values for the SMB beta $s$ and the HML beta $h$. As we move along the size dimension from group A to E, the estimated numbers for $s$ move from positive to negative. Likewise, as we move along the value dimension, from group 1 to 5, the estimated numbers for $h$ move from negative to positive. It tells us that indeed there is commonality in movement among small stocks that is different from large stocks. Regressing returns of small stocks on the SMB factor picks up this comovement. Similarly, values stocks comove together in ways that are different from growth stocks. Hence the HML factor. Overall, these regression results tell us that

size and value are not simply characteristics. Putting small stocks together against big stocks actually forms a factor. Likewise, putting value stocks together against growth stocks forms another factor.

Comparing the R-squared numbers in the one-factor regression with the three-factor regression also tells a similar story. For example, the R-squared's for small stocks (in Groups A) are around 60% in the one-factor regression. In the three-factor regression, the R-squared's increase to around 90%. Of course, having two more factors always improve the R-squared, but not by this much. This is telling us that the SMB factor is picking additional commonality in small stocks.

- **The Fama-French three-factor model:** Borrowing from the CAPM, the pricing relation of the three-factor model is pretty straightforward:

$$E(R_t^i) - r_f = \beta_i \left(E\left(R_t^M\right) - r_f\right) + s_i\, E\left(R_t^{\text{SMB}}\right) + h_i\, E\left(R_t^{\text{HML}}\right) .$$

The risk premiums for the three factors can be estimated using the historical data. Overall, the size premium is somewhat weak in recent periods. The estimated size premium is 3.20% with a t-stat of 1.68. So it is not really significant. The value premium is stronger: 5.15% with a t-stat of 2.78. For the same sample period from 1962 to 2014, the market risk premium is 6.46% with a t-stat of 2.64.

The empirical performance of the Fama-French three-factor model is plotted in Figure 2. Comparing this plot against the one for the CAPM, we can see a clear improvement. By now, we are not surprised that it would work. In the three-factor model, small stocks have a positive factor loading on SMB and are compensated for this exposure. So the model-predicted risk premium is higher than that in the CAPM, where only beta matters. Likewise, value stocks have a positive factor loading on HML and are compensated for this exposure. As a result, in Figure 2, the dots for those portfolios in groups 4 and 5 move horizontally to the right, while those in group 1 move horizontally to the left. So effectively, by having the two added dimensions along size and value, the model performs better.

- **Use the Fama-French three factor model:**

The three-factor model can be used as a benchmark model to evaluate the performance of fund managers. For example, you can put Peter Lynch's performance on the left hand side and regress it against the three factors. You can investigate his exposures to the factors and evaluate how much of his performance derives from such exposures.


Figure 2: The empirical performance of the Fama-French three-factor model, using the Fama-French 25 portfolios. For each portfolio $i$, its risk premium measured from the data is plotted against that predicted by the model.

The alpha from the regression tells you the magnitude of his performance that cannot be explain by the three factors.

A fund manager might have a pretty nice looking CAPM alpha, but when evaluate his performance against the three factor model, his three-factor alpha might be insignificant. This implies that most of his CAPM alpha in fact comes from exposures to the size or value factor. This is what people mean when they say "beta in disguise." For this fund manager, his CAPM alpha actually comes from a beta exposure to a previously unknown risk factor called size or value. Maybe this is why as this quant investing approach moves into the world of mutual funds and ETFs, people are not selling them as alpha's anymore. Instead, they are emphasizing on beta and risk factors.

# Appendix

## A On Running Multivariate Regression

Many students do not like the fact that SMB and HML are not orthogonal to the market portfolio. For example, using annual data from 1962 to 2014, let's regress SMB on the market:

$$R_t^{SMB} = \alpha^{\text{SMB}} + \beta^{\text{SMB}} \left(R_t^M - r_f\right) + \epsilon_t \,.$$

We have a CAPM beta of 0.22, indicating that small stocks on average have a slightly higher beta than large stocks. Run the same regression using the HML factor:

$$R_t^{HML} = \alpha^{\text{HML}} + \beta^{\text{HML}} \left(R_t^M - r_f\right) + \epsilon_t \,.$$

you get a CAPM beta of -0.21, indicating that growth stocks on average have a slightly higher beta than value stocks.

How does this affect our multivariate regression in Equation (2)? The only real effect is that the beta in the three-factor regression is no longer the CAPM beta. Other than this, I cannot think of any significant "damage" of having a factor that is slightly correlated with the market. Of course, Fama and French form their factors using this long/short strategy exactly to take out the market component. As in many situations, simplicity is preferred. In this case, it is really more simple and intuitive to use SMB and HML. If the cost is not being able to read the CAPM beta directly from the three-factor regression, then it is an acceptable cost.

Recall that we call $E(R^{\text{SMB}})$ and $E(R^{\text{HML}})$ the value and size premiums. If we want to be really careful, we should call them the average returns of SMB and HML. The alpha of the above regression, $\alpha^{\text{SMB}}$ gives us the true performance of SMB: 1.76% with a t-stat of 0.91. And $\alpha^{\text{HML}}$ is 6.51% with a t-stat of 3.44. So indeed, the size premium is small and insignificant for the period from 1962 to 2014, while the value premium is pretty strong.

Also, in the above regressions, you will never put $R_t^{\text{SMB}} - r_f$ on the left hand size. This is because $R_t^{\text{SMB}}$ is already a long/short portfolio. If you really want, you could do

$$R_t^{\text{small}} - r_f = \alpha^{\text{small}} + \beta^{\text{small}} \left(R_t^M - r_f\right) + \epsilon_t \,,$$

or

$$R_t^{\text{big}} - r_f = \alpha^{\text{big}} + \beta^{\text{big}} \left(R_t^M - r_f\right) + \epsilon_t \,.$$

Moreover, you notice that $\alpha^{\text{SMB}} = \alpha^{\text{small}} - \alpha^{\text{big}}$ and similarly for beta.

# B Matlab Code

Code 1: Test the models using the Fama-French 25 portfolios

```
n_model=input('which model? Market (1); FF 3 Factor (2)');

load FF_Factors.txt;
start_time=196201;
end_time=201606;
FF_Factors=FF_Factors(FF_Factors(:,1)>=start_time & FF_Factors(:,1)<=
    end_time,:);
Market=FF_Factors(:,2)/100;
SMB=FF_Factors(:,3)/100;
HML=FF_Factors(:,4)/100;
RF=FF_Factors(:,5)/100;

switch n_model,
  case 1, X=Market;
  case 2, X=[Market SMB HML];
end

load FF_Portfolio_25.txt;
FF_Portfolio_25=FF_Portfolio_25(FF_Portfolio_25(:,1)>=start_time &
    FF_Portfolio_25(:,1)<=end_time,:);
n_Portfolio=size(FF_Portfolio_25,2)-1;
Portfolio=FF_Portfolio_25(:,2:end)/100-kron(RF,ones(1,n_Portfolio));
Name=['A1';'A2';'A3';'A4';'A5'; ...
      'B1';'B3';'B3';'B4';'B5'; ...
      'C1';'C2';'C3';'C4';'C5'; ...
      'D1';'D2';'D3';'D4';'D5'; ...
      'E1';'E2';'E3';'E4';'E5'];

% output for alpha, beta, and R2 tables
if n_model==1, beta_CAPM=[]; alpha_CAPM=[]; R2_CAPM=[]; end
if n_model==2, beta_FF3=[]; alpha_FF3=[]; R2_FF3=[]; end
for i=1:n_Portfolio,
```

```
    [b,R2]=Reg_OLS(Portfolio(:,i),X);
    switch n_model,
    case 1,
       R2_CAPM=[R2_CAPM; R2*100];
       beta_CAPM=[beta_CAPM; b(1,2:end)];
       alpha_CAPM=[alpha_CAPM; [b(1,1)*12*100 b(3,1)]];
    case 2,
       R2_FF3=[R2_FF3; R2*100];
       beta_FF3=[beta_FF3; b(1,2:end)];
       alpha_FF3=[alpha_FF3; [b(1,1)*12*100 b(3,1)]];
    end
end

if n_model==1, beta_out=beta_CAPM; else, beta_out=beta_FF3; end;
Y=mean(Portfolio)';
Y_fitted=beta_out*mean(X)';

figure(n_model); plot(Y_fitted*12,Y*12,'r.')
switch n_model,
case 1, title('\bf The Empirical Performance of the CAPM');
case 2, title('\bf The Empirical Performance of the Fama-French Three Factor
     Model');
end
axis([0.02 0.14 0.02 0.14])
hold on;
for k=1:n_Portfolio
     text(Y_fitted(k)*12+0.001,Y(k)*12,[' ' char(Name(k,1))]);
     text(Y_fitted(k)*12+0.004,Y(k)*12,[' ' char(Name(k,2))]);
     if n_model == 1,
       if k==5, text(0.080,Y(k)*12,'Small Value'); arrow([0.080,Y(k)
           *12],[0.072,Y(k)*12]);end
       if k==1, text(0.10,Y(k)*12,'Small Growth'); arrow([0.10,Y(k)
           *12],[0.092,Y(k)*12]);end
       if k==25, text(0.023,Y(k)*12,'Big Value'); arrow([0.043, Y(k)
           *12],[0.052,Y(k)*12]); end
       if k==21, text(0.045,0.040,'Big Growth'); arrow([0.055 0.043],[0.06,Y(
           k)*12]); end
     else
```

```
        if k==1, text(0.108,Y(k)*12,'Small Growth'); arrow([0.105,Y(k)
            *12],[0.095,Y(k)*12]);end
        if k==5, text(0.083,Y(k)*12,'Small Value'); arrow([0.105,Y(k)
            *12],[0.115,Y(k)*12]);end
        if k==25, text(0.118,Y(k)*12,'Big Value'); arrow([0.116, Y(k)
            *12],[0.106,Y(k)*12]); end
        if k==21, text(0.025,0.07,'Big Growth'); arrow([0.038 0.067],[0.038,Y(
            k)*12+0.003]); end
    end
end
if n_model == 1,
 xlabel(['\bf Predicted by the CAPM: \beta^i \times \lambda^M'] );
else,
 xlabel(['\bf Predicted by the FF model'] );
end
ylabel('\bf Measured from the data')
hold on
plot([0.02 0.16],[0.02 0.16],'b--')
hold off
```

Code 2: My OLS Regression Function

```
function [out,R2]=Reg_OLS(Y,X)

A=[ones(length(Y),1) X];
b=inv(A'*A)*(A'*Y);
Eps=Y-A*b;
SE=sqrt(diag(inv(A'*A)*var(Eps)));

out=[b'; SE'; (b./SE)'];
R2=1-var(Eps)/var(Eps);
% use this if need adjusted R2: adj_R2=R2-(1-R2)*size(X,2)/(size(X
   ,1)-size(X,2)-1);
```

Jun Pan
junpan@mit.edu

253-3083
E62-624

## Classes 6 & 7: Equity in the Cross Section, Part 2

This Version: September 26, 2016

# 1 The Momentum Profit and the Four-Factor Model

- **Momentum, past and present:** The momentum profit is the strangest thing. You sort stocks by their past returns into past winners and past losers. In the next few months, the winner portfolio keeps "winning" and the loser portfolio keeps "losing." I can imagine the initial reaction received by Prof. Jegadeesh and Titman when they first presented their results: not very warm. If I were there, I would have asked: could there be a coding error?

  Things certainly have changed since 1993, when the momentum paper was first published in the *Journal of Finance*. Now there is momentum everywhere. Since the late 1990s, hedge funds have been doing long/short momentum strategies in equity, international equity, commodity futures, and others. Since the late 2000s, momentum-style equity mutual funds are being offered to "regular" investors; and now you can also buy momentum factor ETFs. Of course, in the world of mutual funds and ETFs, you can only take long positions in the past winners. As a result, the number one risk exposure in these products remains to be the market risk, not momentum.

- **Forming momentum portfolios** The momentum strategy itself is very simple and the exact portfolio formation varies. By now, the strategy adopted by most fund managers is: in month t, sort stocks by their month t-12 to month t-2 cumulative returns. Notice that the returns in month t-1 are intentionally left out. It is well known that, over the one-week up to one-month horizon, stock returns exhibit reversals (the also famous short-term reversal). So including the month t-1 returns would contaminate the momentum signal.

  As usual, let's double sort by size and momentum to get 25 (5x5) portfolios. As shown in Table 1, each portfolio is indexed by size (A to E) and momentum (1 to 5). Our focus is on the momentum dimension, but size is always an important control variable

Table 1: Momentum Portfolios in the CAPM and the Fama-French Three-Factor Model. The 25 portfolios are double sorted by size (from A to E) and past returns (from 1 to 5). All $\alpha$'s are reported in annualized terms (x12). Statistically significant $\alpha$'s are reported in bold. Monthly data from January 1962 to July 2015.

| | CAPM | | | | The FF Three Factor Model | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Portfolio | $\alpha$ (%) | t-stat | $\beta$ | R2 (%) | $\alpha$ (%) | t-stat | $\beta$ | $s$ | $h$ | R2 (%) |
| A1 | **-8.19** | -3.31 | 1.37 | 57.99 | **-12.14** | -6.75 | 1.19 | 1.24 | 0.41 | 78.50 |
| A2 | 1.68 | 1.00 | 1.05 | 63.57 | **-2.46** | -2.66 | 0.94 | 0.97 | 0.52 | 89.40 |
| A3 | **5.01** | 3.33 | 0.99 | 66.03 | 1.21 | 1.56 | 0.89 | 0.89 | 0.48 | 91.27 |
| A4 | **6.57** | 4.36 | 1.00 | 66.72 | **3.39** | 4.32 | 0.88 | 0.92 | 0.35 | 91.26 |
| A5 | **8.87** | 4.64 | 1.21 | 64.28 | **6.84** | 6.20 | 0.98 | 1.14 | 0.09 | 88.49 |
| B1 | **-7.25** | -3.44 | 1.45 | 68.20 | **-10.27** | -6.18 | 1.31 | 0.95 | 0.32 | 80.84 |
| B3 | 0.95 | 0.65 | 1.12 | 72.39 | **-2.38** | -2.47 | 1.04 | 0.76 | 0.42 | 88.49 |
| B3 | **3.47** | 2.82 | 1.03 | 76.06 | 0.44 | 0.60 | 0.96 | 0.67 | 0.39 | 91.81 |
| B4 | **5.69** | 4.54 | 1.05 | 75.98 | **2.92** | 4.34 | 0.95 | 0.75 | 0.32 | 93.31 |
| B5 | **6.97** | 4.16 | 1.28 | 72.38 | **5.97** | 5.82 | 1.06 | 0.95 | -0.05 | 89.99 |
| C1 | **-5.54** | -2.78 | 1.37 | 68.03 | **-7.86** | -4.33 | 1.29 | 0.61 | 0.27 | 74.17 |
| C2 | 0.55 | 0.46 | 1.10 | 78.67 | **-2.13** | -2.19 | 1.07 | 0.46 | 0.38 | 86.76 |
| C3 | **2.34** | 2.18 | 1.01 | 80.01 | -0.45 | -0.59 | 0.99 | 0.46 | 0.40 | 89.98 |
| C4 | **3.19** | 3.08 | 1.01 | 80.94 | 0.77 | 0.97 | 0.98 | 0.43 | 0.34 | 89.31 |
| C5 | **6.87** | 4.58 | 1.21 | 74.71 | **6.51** | 5.80 | 1.04 | 0.70 | -0.11 | 86.23 |
| D1 | **-6.11** | -3.08 | 1.34 | 67.05 | **-8.24** | -4.24 | 1.33 | 0.31 | 0.31 | 69.53 |
| D2 | -0.05 | -0.04 | 1.11 | 79.82 | **-2.25** | -2.06 | 1.14 | 0.17 | 0.36 | 83.30 |
| D3 | 1.83 | 1.98 | 1.00 | 84.15 | -0.29 | -0.36 | 1.03 | 0.16 | 0.35 | 88.30 |
| D4 | **3.59** | 4.26 | 0.99 | 86.29 | **2.10** | 2.69 | 1.01 | 0.15 | 0.23 | 88.57 |
| D5 | **5.49** | 4.03 | 1.15 | 76.12 | **5.52** | 4.55 | 1.03 | 0.44 | -0.12 | 81.64 |
| E1 | **-5.79** | -3.07 | 1.24 | 65.92 | **-6.68** | -3.54 | 1.30 | -0.13 | 0.20 | 66.96 |
| E2 | -0.33 | -0.28 | 0.94 | 73.42 | -1.28 | -1.12 | 1.03 | -0.20 | 0.22 | 76.88 |
| E3 | -0.88 | -1.08 | 0.90 | 84.74 | -1.41 | -1.90 | 0.98 | -0.20 | 0.15 | 87.82 |
| E4 | 1.20 | 1.46 | 0.89 | 84.10 | 1.19 | 1.57 | 0.95 | -0.23 | 0.06 | 86.85 |
| E5 | **3.30** | 2.70 | 1.02 | 75.83 | **4.47** | 3.69 | 0.99 | -0.04 | -0.21 | 76.95 |

in any trading strategy. Ideally, we would like a strategy to work within each size group, from A to E. And as shown in Table 1, the momentum strategy delivers such a result. It should be noted that I use monthly returns to run the regressions, but report the alpha's in annualized terms for ease of communication (which might be a source of confusion by now). In any case, the reported alpha's in Table 1 are the monthly alpha's multiplied by 12. All other estimates are unaffected.

- **The momentum profit:** By now, I believe that you know how to read and evaluate the numbers in Table 1. So I'll be brief.

  Focusing on one size category, say group A, and varying from A1 to A5, we move from portfolios containing past losers to past winners. You can see the strong magnitudes of these alpha's and their t-stat's: the CAPM alpha is -8.19% per year for A1 and 8.87% for A5. Both estimates are statistically significant with large t-stat's. It is also nice that within each size group, the alpha increases monotonically from group 1 to 5. Moreover, even for stocks in the large cap group, the momentum profit is quite strong and statistically significant: the CAPM alpha is -5.79% for E1 and 3.30% for E5. Recall that for book-to-market, the results are not this strong for group E. Moving to the right side of the Table, we see that the Fama French factors do not help us explain the momentum profit. Not at all.

- **More observations:** By now, some of you might have come to like reading numbers. If so, you could spend even more time on Table 1. Notice how the CAPM beta's for the two extreme portfolios (winner and loser) tend to be larger than the middle three portfolios? This indicates that momentum portfolios tend to be more volatile. Of course, if you are doing the long/short strategy, then the beta exposure decreases to a large extend. Still the momentum strategy tend to be more volatile compared with other strategies (see also page 3 of Prof. Kent Daniel's slides where he reports the standard deviations of the six popular strategies pursued by GSAM.)

  Focusing on the size and value exposures in Table 1, you might also notice that the winner portfolios tend to have negative exposures to HML while the loser portfolios tend to have large and positive exposures to HML. This tells you that there is an interaction between these two signals: growth stocks tend to be past winners or past winners tend to be growth stocks. So to sharpen your momentum signal, you might want to take advantage of this interaction term: hold past winners with high book-to-market ratio and sell past lowers with low book-to-market ratio.

- **Paper alpha vs. real alpha:** While the momentum profit looks impressive on paper,

the real alpha of the trading strategy might not be as impressive because of the execution costs involved with high portfolio turnovers. For example, the annual turnover of a small-cap momentum mutual fund is close to 200%. So the real alpha of the strategy will be cut by transaction costs. One of the main sources of transaction costs is price impact, especially for a large fund pursuing momentum strategy in small-cap stocks, where liquidity is known to be poor.

In general, keeping the execution costs low should be as important as generating alpha. Low execution costs contribute directly to portfolio performance. In today's trading environment, knowing how to trade large institutional-size portfolios to minimize transaction costs separates a good asset manager from a mediocre one.

- **Momentum in mutual funds and ETFs:** For long-only equity mutual funds or ETF pursuing momentum strategies, the typical momentum portfolio contains stocks that are ranked by past performance among the top 1/3. For example, if we focus on large-cap stocks in groups E, then the momentum portfolio is a value-weighted portfolio of all the stocks in our E5 plus the top half of E4. For small-cap momentum funds, it is a value-weighted portfolios of all the stocks in our A5 and the top half of A4. As you can see in Table 1, such portfolios do have positive alphas. At the same time, however, they also have a pretty large exposure to the market risk. In other words, by holding a momentum portfolio, an investor's number one risk exposure is not really momentum. By pushing quant investing into the long-only space, the razor-sharp focus on the targeted risk is lost because of the inability to do long/short.

- **The four-factor model:** Because the momentum profit cannot be explained by the Fama-French factors, we add the momentum factor to the FF three-factor model to form a four-factor model:

$$E(R_t^i) - r_f = \beta_i \left(E\left(R_t^M\right) - r_f\right) + s_i\, E\left(R_t^{\text{SMB}}\right) + h_i\, E\left(R_t^{\text{HML}}\right) + w_i\, E\left(R_t^{\text{MOM}}\right) ,$$

where the new MOM factor is constructed in a way similar to the HML factor. Along the size dimension, stocks are sort into two groups: small or big. Along the momentum dimension, stocks are sort into three groups with 30% in high past returns, 40% in neutral, and 30% in low past returns. Again, because these portfolios are to be used to construct factors, one would like to have them as diversified as possible. Hence the coarser sort. Using these portfolios, the momentum factor is constructed as,

$$R^{\text{MOM}} = R^{\text{winner}} - R^{\text{loser}} ,$$

where $R^{\text{winner}}$=1/2 (small high + big high) and $R^{\text{loser}}$=1/2 (small low + big low). Finally, the factor exposures ($\beta$, $s$, $h$, $w$), can be estimated using the four-factor regression:

$$R_t^i - r_f = \alpha_i + \beta_i \left(R_t^M - r_f\right) + s_i \, R_t^{\text{SMB}} + h_i \, R^{\text{HML}} + w_i \, R^{\text{MOM}} + \epsilon_t^i \, .$$

This four-factor model is also call the Carhart model, because it was first proposed in a 1997 Journal of Finance paper written by Mark Carhart to examine the performance of equity mutual funds. Carhart was a PhD student of Prof. Fama and helped run the Global Alpha fund, founded by Cliff Asness in the late 1990s, at GSAM. At its peak, the team managed over $185 billion in assets. In 2011, the fund was closed by Goldman. It marked the end of an era.

- **Using the four-factor model:** The four-factor model can be used as a benchmark model to evaluate the performance of fund managers. A fund manager following momentum strategy might have a pretty nice looking FF3 alpha, but when evaluate his performance against the four factor model, his four-factor alpha might be insignificant. This implies that most of his FF3 alpha comes from exposures to the momentum factor. Again, "beta in disguise."

# 2 Quant Investing: crowded trades, over-used signals

- **Popular quant signals:** By now, there is a set of well established quant signals, with size, value, and momentum being the most basic collection. For example, in Prof Kent Daniel's slides, he mentioned six quant-style portfolios held by GSAM's Global Equity Opportunities funds around 2007.

  Value and momentum are two of the six quant signals. In addition, profitability, measured as the earnings-to-sales ratio, is also a useful quant signal. Recently, Fama and French propose a profitability signal that uses the operating profit (=revenues minus cost of goods sold, minus selling, general, and administrative expenses, minus interest expense) divided by book value of equity. They find that stocks with robust profitability overperforms stocks with weak profitability and create a factor called RMW (robust minus weak).

  By now, you might notice that accounting data plays a pretty important role in signal creation. This indeed is true. Many of the quant signals were first reported by account-

ing professors. The quant signals relating to earnings quality is one such example. In a 1996 paper published in the *Accounting Review*, Prof Sloan shows stock prices do not fully reflect the information in the quality of earnings. Stocks with low earnings quality (high accrual) underperform stocks with high earnings quality.

The quant signal using analysts forecast revision also comes from the accounting literature. In a 1991 paper, also published in the *Accounting Review*, Prof Stickel finds that analysts revision affect prices, but prices do not immediately assimilate the information. In fact, prices continue to drift in the direction of the revision for about six months after the revision. Another important pattern related to earnings news is reported by Bernard and Thomas (1989). This is the famous post earnings announcement drift: stocks with positive earnings surprises on their announcement day keep drift upward in their stock prices a few weeks (up to 60 days) after the announcement while stocks with negative earnings surprises keep drifting downward.

Finally, the sixth signal reported in Prof Kent Daniel's slides is management impact. This signal builds on two observations. Loughran and Ritter (1994) reports long-term underperformance after IPO or SEO (seasoned equity offering), and Ikenberry, Lakonishok, and Vermaelan (1995) reports long-term overperformance after announcements of share repurchases. In their recent paper, Fama and French introduces a signal that is similar in spirit. They use the firm's asset growth as a signal for firm investment. They find that stocks with low investment (low asset growth) outperform stocks with high investment. Calling firms with low investment conservative, and high investment aggressive, Fama and French introduce a new factor called CMA (conservative minus aggressive). Together with the market portfolio, SMB, HML, RMW (just mentioned), Fama and French build a new five-factor model.

- **Crowded trades and over-used signals:** By now, popular quant signals are a common knowledge. This is an over-crowded space with over-used signals. Moreover, the transparency of these trading strategies also makes the funds easy to predict, inviting front runners.

  The 2007 quant meltdown is clearly a result of over-crowding. This example itself is interesting, but the lesson to be learned is not confined just to this one space. To a large extent, the 1998 LTCM crisis was a parallel example in the fixed income arbitrage space. Since the mid-1980s, the fixed-income market has enjoyed a great bull run with an overall trend of decreasing interest rates (from double digits). By the early 1990s, many fixed income arbitrage funds are having a lot of success. Success breeds imitation.

As a result, the market became over-crowded with many hedge funds in the space of fixed-income arbitrage, doing similar yield curve trading. Sounds familiar?

In the case of LTCM, the actual trigger was Russia's default on its local currency debt, which LTCM did not have a lot of exposure to. Similarly, the initial trigger for the 2007 quant meltdown was disruptions in the sub-prime mortgage market, which most of the quant funds did not have any direct holdings. The sub-prime disruption later spilled over to the credit market, and to currency carry trades. At the time, what many quant investors didn't realize was that the success in their space attracted participation from investors outside of the quant space: statistical arbitrage and other multi-strategy hedge funds.

As the multi-strategy hedge funds experienced the disruptions in the other markets, they sought to liquidate assets to raise more cash. The least costly and the quickest approach is to liquidate the most liquid holdings, which are the stocks in their quant strategies. Hence the typical contagion story. The quant stocks started to spiral down together not because they shared some negative fundamentals. Instead, the co-movement was caused by the commonality in who were holding these stocks: quant funds. For the 2007 quant meltdown, you need a special pair of quant goggles to see it. Otherwise, the market looked quite normal during the first two weeks of August 2007. But in the quant world, the portfolios were moving down by as much as 20 sigmas. The draw-down affected all quant strategies in all geographical regions.

Similarly, back in 1998, the wall street firms had all the incentive to save LTCM because they were holding similar assets and pursue the same trading strategies as LTCM. If LTCM liquidated their portfolios to the market, the liquidity crisis will bring down many of the investment banks. You might wonder: a liquidity crisis is only temporary, why worry? Everything will bounce back, right? For example, in the second half of August 2007, the quant funds rebounded and things were back to normal. Well, if you are holding a leveraged position, then this is a totally different story: you might not be able to survive the temporary liquidity crisis. Being levered during a liquidity crisis brings into my mind the picture painted by one of Warren Buffett's famous quotes: Only when the tide goes out do you discover who's been swimming naked.

In the case of LTCM, the leverage of of the fund has been widely documented and the often quoted number is 30 to 1. (See, for example, the book by Roger Lowenstein). In the case of the quant meltdown, Bob Litterman gave this description of GSAM's Global Equity Opportunities fund. There were +1000 positions on individual stocks, with an average holding period in months. The portfolio is market neutral and industry

neutral, with a volatility of about 10% per year and 1.4% per week. Up to 2007, the average return was about 15% per year. In July 2007, however, it was down 15%. The overall size of the fund was about $6 billions, with $24 billion long/short positions. So effectively, with $6 billion equity, the firm's assets were at around $48: 8 to 1. From August 1 to 10, the fund was down 30%, an over 20 sigma drawdown.

- **What next?** The lessons learned from the quant meltdown:
  - cannot be too big: whale.
  - cannot be too crowded: run for the exit.
  - cannot be too transparent: front running.

  Clearly, it is important to have your unique trading strategies. As such, the search for new quant signals is still on. Given the massive amount of "data mining" in the past ten to twenty years, the amount of interesting signals left for us to discover might be limited. Overall, this area is just not as exciting and creative as it was ten or twenty years ago.

  An alpha that looks good on paper does not necessarily translate to real alpha. Transaction costs such as price impact or short-sale constraint cut into the real alpha. This is especially true for smaller and less liquid stocks. Unfortunately, most of the quant signals work better in small to medium stocks. Another problem is that some quant signals that used to work in the past ceased to work after the publication of the signal.

  One push is to other asset classes, such as fixed income. But the fixed-income world is probably smarter and faster than the equity world in the sense that most of the fixed income arbitrage trades are indeed designed to exploit cross-sectional pricing differences. For the corporate bond market, the lack of liquidity in that market does not make it a suitable place for the traditional quant investing.

  Another recent push is to mutual funds and ETFs. As we discussed earlier, for long-only space, a large portfolio of the risk exposure comes not from the quant signal, but from the market risk. This probably is the most limiting aspect of quant investing in this space. Nevertheless, the push in that direction seems to be a recent trend. In yesterday's *Financial Times*, it was reported that Goldman has also joined the "smart beta" ETF rush.

# 3 Currency Carry Trade

- **The FX market:** In terms of trading, the foreign exchange market is the largest and the most liquid market in the world. According to a BIS survey in 2013, the daily trading volume of the currency market was $5.34 trillion, among which the dollar trading volume in the spot market was around $2 trillion. By comparison, the average daily dollar trading volume of NYSE group is $41 billion in 2015. The US Treasury market, important and highly liquid, has a daily dollar trading volume around $500 billion. There are mixed trading motives behind the trillion dollar daily trading volume: hedging currency exposure could be an important component, but currency speculation accounts for a sizable percentage of the trading volume.

- **Currency carry trade:** Currency carry trade is one of the well known trading strategies pursued by macro hedge funds. In a way, it is like quant investing, except that the history of this trading strategy is probably longer than quant investing in the equity space. Fama (1984) was one of the earlier papers documenting this pattern, which is called forward premium puzzle in academic.

  The strategy is simple and intuitive. Currencies with high interest rates (e.g., New Zealand dollar or Australia dollars) are used as asset or target currencies and currencies with low interest rates (Japanese Yen) are used as funding currencies. The strategy is to buy the "target" currencies and borrow from the "funding" currency, carry this position with a positive carry, and unwind it later in the spot market: sell the target currency and buy back the funding currency. As a result, there are two drivers for the portfolio returns: the interest rate differential, and the gain or loss in the spot market when unwinding the trade. It is very similar to the two components in stock returns: dividend yield and capital gains.

  On average, this is a profitable trading strategy, but it is sensitive to the liquidity condition of the global markets. Large losses in currency carry often incurs when there is a global sell-off of risky assets. In a flight to quality, investors typically abandon the risky assets and move to the safer securities such as US treasury or the perceived safe haven currencies (e.g., the Swiss Franc, the Japanese Yen, and the US dollars). Accompanied with the large losses in currency carry is the sudden strengthening of the funding currency and weakening of the target currency. As carry traders rush to the market to unwind their carry trades, the situation is further exacerbated.

- **Currency Carry Profit:** Let's apply the portfolio approach we've learned from the

quant investing to currency carry. Let's use the US dollar as an anchor and calculate portfolio returns from the perspective of a US investor. In month t, he borrows in US dollar and buys one specific foreign currency. In month t+1, he unwinds the trade and calculates the realized return.

In forming the portfolios, we use the interest rate differential between the foreign and US one-month risk-free rates, $i^* - i$ as the quant signal. We sort foreign currencies into six groups. Group 1 contains currencies with the highest interest rates: the target currencies. Group 6 contains currencies with the lowest interest rates: the funding currencies. For each portfolio, we calculate the holding period returns from each currency and equal weight the returns across the currencies in the portfolio.

Table 2: Currency Portfolios Sorted by Interest Rates

| | | CAPM | |
|---|---|---|---|
| Rank | exret (%) | beta | alpha (%) |
| 1 | 0.79 | 0.19 | 0.69 |
| | [4.56] | [3.08] | [3.22] |
| 2 | 0.35 | 0.17 | 0.26 |
| | [2.39] | [3.64] | [1.55] |
| 3 | 0.28 | 0.12 | 0.22 |
| | [2.14] | [2.36] | [1.39] |
| 4 | 0.15 | 0.08 | 0.11 |
| | [1.21] | [1.91] | [0.77] |
| 5 | -0.05 | 0.07 | -0.08 |
| | [-0.38] | [1.53] | [-0.58] |
| 6 | -0.18 | 0.01 | -0.18 |
| | [-1.37] | [0.24] | [-1.30] |

Table 2 uses monthly data from January 1987 through December 2011. The number of available currencies varies over time. For the period from 1987 through 2011, the sample starts with 17 currencies and reaches a maximum of 34 currencies. Since the launch of Euro in January 1999, the sample covers 24 currencies.

The average excess return for portfolio 1 is 0.79% per month and is statistically significant. By comparison, the average excess return for portfolio 6 is slightly negative and is insignificant. A typical currency carry trade would long portfolio 1 and short portfolio 6. The difference between these two portfolio returns constitutes the typical currency carry profit: around 0.97% per month (roughly 11% per year). Using the US stock market portfolio as a benchmark, we find that the portfolio of target currencies has a beta of 0.19, which is interesting given that the portfolio involves positions on

currencies with no direct exposure to the stock market. By contrast, the portfolio of funding currency has very little exposure to the US stock market. Overall, the CAPM alpha of the currency carry trade remains large and significant.

Jun Pan
junpan@mit.edu
MIT Sloan School of Management
15.433-15.4331 Financial Markets, Fall 2016
253-3083
E62-624

## Classes 9 & 10: Equity in the Time Series, Part 2

## Time-Varying Volatility

This Version: September 30, 2016

Just when Prof. Fama and his PhD/MBA students were busy working on the cross section of expected stock returns, another area of Finance was taking shape. In this area, tools developed in Econometrics and Statistics are applied to financial time series such as the time-series of stock returns. Given how difficult it is to estimate the first moment (expected returns), much of the attention was devoted to estimating the second moment, stock return volatility. The most visible figure in this area is Prof. Rob Engle, who was awarded a Nobel Prize in 2003 for "methods of analyzing economic time series with time-varying volatility (ARCH)." The ARCH paper was published in 1982 when Prof. Engle was an Economics professor at UCSD. The more famous GARCH extension was later published in 1986 by his PhD student Prof. Bollerslev.

# 1 Volatility models and market risk measurement

- **The need for better risk management tools:** In the early 1990s, there were two developments that made volatility models attractive and relevant. First, the need of a better option pricing model becomes quite obvious after the 1987 stock market crash. The Black-Scholes model builds a very good foundation, but it lacks flexibility in handling the richer reality. In the Black-Scholes model, stock returns are normally distributed with a constant volatility $\sigma$. Any casual inspection of the data would inform us that volatility is not a constant. So having a better volatility model would be a first step toward a better option pricing model.

  Second, and even more pressing was the need for better risk management tools. The mortgage-backed security was developed in the 1980s, and the over-the-counter derivatives market started to take off by the late 1980s. By the early 1990s, the increasing activity in securitization and the increasing complexity in the fixed-income products have made the trading books of many investment banks too complex and diverse for

the chief executives to understand the overall risk of their firms.

It was an industry wide phenomenon. For example, in "Money and Power," Cohan wrote about the difficult year of 1994 at Goldman after the two amazingly profitable years in fixed-income in 1992 and 1993. In the book, he quoted Henry Paulson, "What came out of the 1994 debacle was best practices in terms of risk management, The quality of the people, and the processes that were put in place – anything from the liquidity management to the way we evaluated risk and really the independence of that function – changed the direction of the firm."

- **JPMorgan's RiskMetrics:** Among all Wall Street firms, JPMorgan's effort was by far the most visible and influential. This 2009 New York Times article titled "Risk Mismanagement" gave a very good account of the events.

  > JPMorgan's chairman at the time VaR took off was a man named Dennis Weatherstone. Weatherstone, who died in 2008 at the age of 77, was a working-class Englishman who acquired the bearing of a patrician during his long career at the bank. He was soft-spoken, polite, self-effacing. At the point at which he took over JPMorgan, it had moved from being purely a commercial bank into one of these new hybrids. Within the bank, Weatherstone had long been known as an expert on risk, especially when he was running the foreign-exchange trading desk. But as chairman, he quickly realized that he understood far less about the firms overall risk than he needed to. Did the risk in JPMorgans stock portfolio cancel out the risk being taken by its bond portfolio – or did it heighten those risks? How could you compare different kinds of derivative risks? What happened to the portfolio when volatility increased or interest rates rose? How did currency fluctuations affect the fixed-income instruments? Weatherstone had no idea what the answers were. He needed a way to compare the risks of those various assets and to understand what his companywide risk was.

  What later became RiskMetrics was an internal effort developed within JPMorgan in 1992 in response to the CEO's question. Quoting the New York Times article again,

  > By the early 1990s, VaR had become such a fixture at JPMorgan that Weatherstone instituted what became known as the 415 report because it was handed out every day at 4:15, just after the market closed. It allowed him to see what every desk's estimated profit and loss was, as compared to its risk,

and how it all added up for the entire firm. True, it didn't take into account Taleb's fat tails, but nobody really expected it to do that. Weatherstone had been a trader himself; he understood both the limits and the value of VaR. It told him things he hadn't known before. He could use it to help him make judgments about whether the firm should take on additional risk or pull back. And that's what he did.

- **Global risk factors:** In 1994, JPMorgan started to make RiskMetrics publicly available. It published its technical document outlining its risk measurement methodologies. It also made available two sets of volatility and correlation data used in the computation of market risk. In Spring 1996, I was hired as a research assistant to Prof. Darrell Duffie to work on risk management and Value-at-Risk. I spent a lot of time reading this technical document of RiskMetrics and I also downloaded the volatility and correlation dataset everyday just to play with them and track the movements. It was through having to deal with the datasets that the immensity of the global financial markets became real to me.

  For example, the 1996 RiskMetrics data files covered over 480 financial time series that were important for the trading books of most investment banks. This includes

  - Equity indices across the world.
  - Foreign exchange rates.
  - The term structure of interest rates across the world:
    * money market rates (1m, 3m, 6m, and 12m) for the short end.
    * government bond zero rates (2y, 3y, 4y, 5y, 7y, 9y, 10y, 15y, 20y, and 30y) for the longer end.
  - The term structure of swap rates across the world (2y, 3y, 4y, 5y, 7y, and 10y).
  - Commodities: spot and futures of varying maturities.

- **The variance-covariance matrix:** In order to measure the firm-wide risk exposure, one need to first map each position to these risk factors, and then calculate the volatility of the overall portfolio or portfolios by asset class: interest rates, equity, currency, and commodities. One of the key building block of this calculation is the variance-covariance matrix of the risk factors. For the 480 risk factors used by JPMorgan in 1996, this involves calculating the volatility for each of the 480 risk factors, and then calculate the pair-wise correlations between the 480 risk factors. In the rest of the class, we will be busying doing these calculations.

Let me quote a few paragraphs from the 2012 annual report of Goldman Sachs so as to give you an update on the risk management effort on Wall Street since 1996.

> We also rely on technology to manage risk effectively. While judgment remains paramount, the speed, comprehensiveness and accuracy of information can materially enhance or hinder effective risk decision making. We mark to market approximately 6 million positions every day. And, we rely on our systems to run stress scenarios across multiple products and regions. In a single day, our systems use roughly 1 million computing hours for risk management calculations.
>
> When calculating VaR, we use historical simulations with full valuation of approximately 70,000 market factors. VaR is calculated at a position level based on simultaneously shocking the relevant market risk factors for that position. We sample from 5 years of historical data to generate the scenarios for our VaR calculation. The historical data is weighted so that the relative importance of the data reduces over time. This gives greater importance to more recent observations and reflects current asset volatilities, which improves the accuracy of our estimates of potential loss. As a result, even if our inventory positions were unchanged, our VaR would increase with increasing market volatility and vice versa.

As you can see, roughly 6 million positions are mapped into 70,000 market factors in Goldman's risk management system. If I understand their statement correctly, this implies a variance-covariance matrix of 70,000 by 70,000.

Back in the mid-1990s, all three of my Chinese classmates at the NYU Physics department went to work at Citibank after graduation. I was the only exception, who went on to get another PhD in Finance. So much for the future of Physics, which was better off without us. During one of my visits back to New York, I visited them at Citibank, thinking how exciting it was for them to be working in the *real* world with a real paycheck. And I was very surprised to see how bored they all looked. One of them worked in the risk management group and his job was to calculate the variance-covariance matrix everyday. He looked miserable. I guess this is not the most exciting job if you have to do it everyday. Many years later, I got an email from Mr. Variance-Covariance, who has become a managing director at Citibank. A happy ending, by Wall Street standard.

# 2 Estimating Volatility using Financial Time Series

In general, volatility is very easy to estimate. Unlike in the case of expected returns, volatility can be measured with better precision using higher frequency data. The convention in this field is to use daily data. For stock market returns, having one month of daily data could get you a pretty accurate estimate. We will continue to use the time series of aggregate stock returns as an example.

I should mention that in this field, log returns are being used more often than percentage returns:

$$R_t = \ln S_t - \ln S_{t-1}\,,$$

where $S_t$ could be date-t stock price, currency rate, interest rate, or commodity futures price. At the daily frequency, the magnitude of returns are generally very small. Using the handy Taylor expansion, we know that for small $x$,

$$\ln(1+x) \approx x\,.$$

Repeating this for log-returns, we have

$$R_t = \ln S_t - \ln S_{t-1} = \ln\left(\frac{S_t}{S_{t-1}}\right) = \ln\left(1 + \frac{S_t - S_{t-1}}{S_{t-1}}\right) \approx \frac{S_t - S_{t-1}}{S_{t-1}}\,.$$

In other words, working with log-returns or percentage-returns does not make too much of a difference when returns are small in magnitude.

Also notice that our attention is no longer on the first moment. Getting the volatility right is our main task. We calculate the variance by,

$$\mathrm{var}(R_t) = E\left(R_t - \mu\right)^2 = E(R_t^2) - \mu^2\,.$$

The volatility estimate is

$$\mathrm{std}(R_t) = \sqrt{\mathrm{var}(R_t)} = \sqrt{E(R_t^2) - \mu^2} = \sqrt{E(R_t^2)} \times \sqrt{1 - \frac{\mu^2}{E(R_t^2)}}\,.$$

At the daily frequency, $\mu$ is around a few basis points for the US equity market, while the daily volatility is around 100 basis points (i.e., 1%). As a result, $\mu^2/E(R_t^2)$ is a really really small number and $\sqrt{1-\mu^2/E(R_t^2)} \approx 1 - \frac{1}{2}\mu^2/E(R_t^2)$ is very close to one. So it does not make a big difference whether or not we subtract $\mu$ from the realized returns such as $R_t$ in

estimating the volatility. You will notice that most of the time, people drop $\mu$ for simplicity:

$$\text{std}(R_t) \approx \sqrt{E(R_t^2)}\,.$$

Of course, this relationship between daily vol and daily $\mu$ exists mostly true for assets in the risky category. For fixed income product, this might not be true. If you would like to be safe, one way to approach the data is to first demean the time-series data: $R_t - \mu$, and then apply the volatility models.

- **SMA:** The simple moving average model fixes a window, say one month, and use the daily returns within this window to calculate the sample standard deviation. The window is then moved forward by one step, say one day, and the whole calculation gets repeated again.

  In Figure 1, I use a moving window of one month and move the window one month at a time to plot a monthly time-series of SMA volatility estimates. In this space, volatility is usually quoted at an annualized level. So I multiply the volatility estimates by $\sqrt{252}$ (assuming 252 business days per year). This annualized volatility corresponds to the volatility coefficient $\sigma$ in the Black-Scholes model. So we also compare these volatility numbers with the option-implied volatility.

  Just so you are convinced that these volatility estimates can be estimated with a pretty good precision using only one month of daily data, I also plotted the 95% confidence intervals. If you compared Figure 1 against Figure 2, you can see the marked difference in estimation precision. Using one month of daily data to estimate the average return, what you get is very much noise.

  Another observation I would like you to make is the variation of market volatility over time. Its pattern is very different from that of market returns. Volatility is persistent: a day of high volatility is usually followed by another day of high volatility. Volatility also tends to spike up once in a while. If you plot these events against the NBER business cycles, you see that volatility usually spikes up during recessions. But recessions are not the only time when volatility spikes up. Whenever the market is in trouble, volatility goes up. Let me name a few recent events: the 1987 stock market crash, the 1997 Asian Crisis, the 1998 LTCM crisis, the 2000-01 tech bubble/burst, the 9/11, and the 2008 financial crisis. Finally, whenever volatility is at an usually high or low level, it tends to revert back to its historical average. This pattern is called mean reversion. Using a longer sample that includes the Great Depression, the historical average of volatility is around 20%. Using the more recent sample, the average volatility is around 15%.

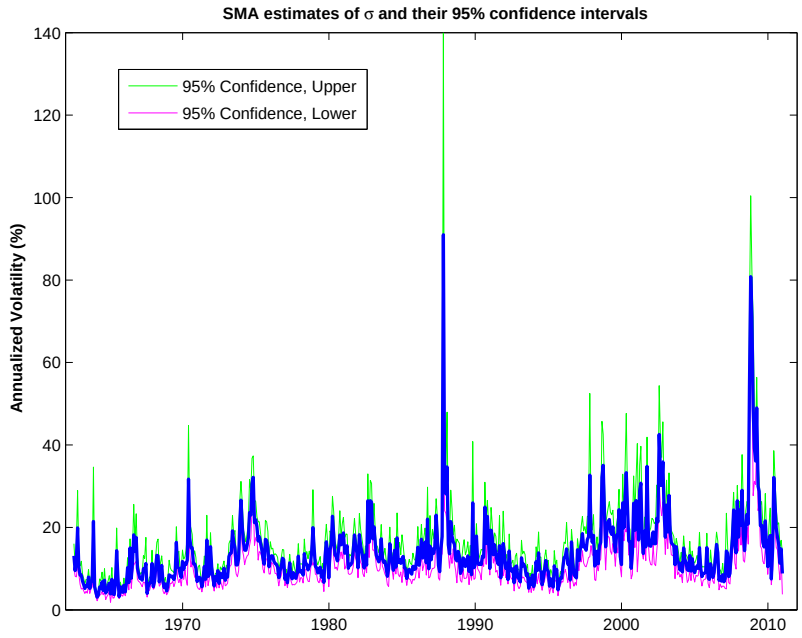


Figure 1: Time-Series of Stock Volatility using SMA.

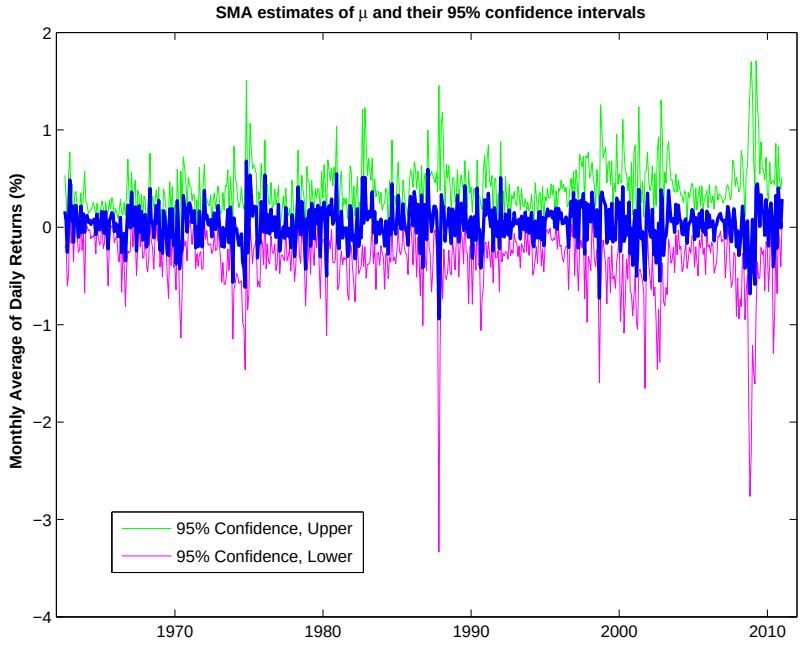


Figure 2: Time-Series of Monthly Average Stock Returns using SMA.

The plot has not been updated for the last few years. Nowadays, the best place to get stock market volatility (without having to do any calculation) is the CBOE VIX index. For example, over the one-week period from August 17 to 24, 2015, the VIX shoots up from 13.02% to 40.74%, because of the concern over the Chinese stock market.

Comparing the volatility estimate with the VIX index, which is effectively the option implied volatility, you notice that the option implied volatility seems to be consistently higher than the volatility estimate. We will visit this issue again when we cover the options market.

- **EWMA:** The exponentially weighted moving average model is an improvement over the SMA model. Instead of applying equal weights to all observations with a fixed window, EWMA applies an exponential weighting schedule. It chooses a decay factor $\lambda$, which is a number between zero and one, and performs the volatility estimate by:

$$\sqrt{(1-\lambda)\sum_{n=0}^{N}\lambda^n\left(R_{t-n}\right)^2}.$$

Let's first put aside the term $1-\lambda$ and focus on the terms within the summation. We put a weight of 1 for today ($n=0$), $\lambda$ for yesterday, $\lambda^2$ for the day before yesterday, and so on. With this weighting schedule, as we move further back into the history and away from today $t$, the contribution of $(R_{t-n})^2$ decreases according to the exponential schedule. Hence the name. Figure 3 gives a graphical presentation of this exponential weighting scheme.

Going back to one of the paragraphs I quoted from Goldman's annual report: "We sample from 5 years of historical data to generate the scenarios for our VaR calculation. The historical data is weighted so that the relative importance of the data reduces over time. This gives greater importance to more recent observations and reflects current asset volatilities, which improves the accuracy of our estimates of potential loss." So effectively, the length of the window $N$ is set at five years and a decay factor is selected to put more weight to the more recent events. In Goldman's report, the value of the decay factor was not reported. In RiskMetrics, $\lambda$ was fixed at 0.94 for all time series. Also, the choice of the window size is not important because the decay factor $\lambda$ effectively selects the window size for you. Figure 3 gives a nice graphical presentation on how the window size is determined by the decay factor: a strong decay factor ($\lambda = 0.8$) implies a smaller window while a mild decay factor ($\lambda = 0.97$) implies a larger window. A window of 5 years is definitely not necessary: the return happened

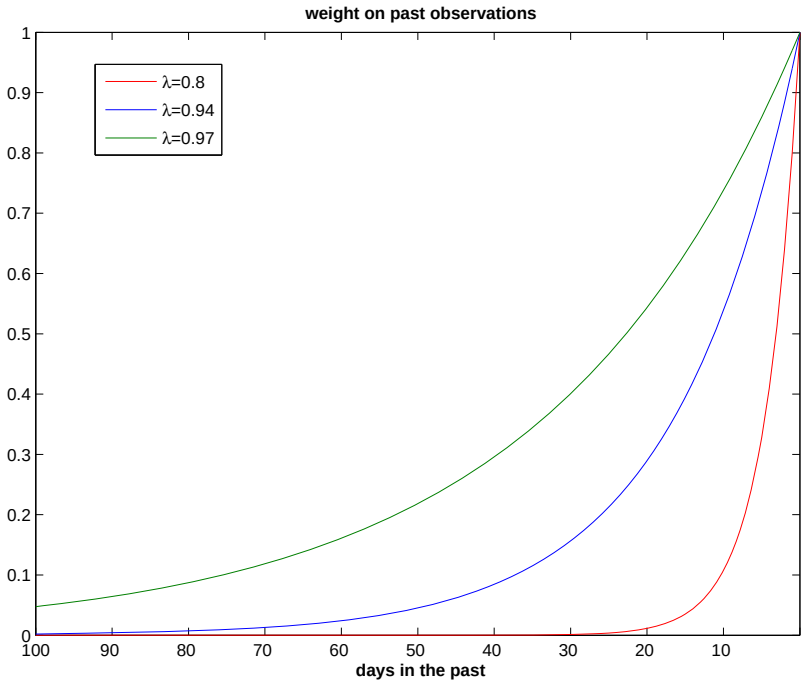


Figure 3: The Exponential Weighting Scheme in EWMA.

5 years ago has a weight of $\lambda^{5\times 252}$, which is too small to matter.

Now let's come back to the term $1-\lambda$. Notice that

$$1+\lambda+\lambda^2+\lambda^3+\ldots=1/(1-\lambda)\,.$$

So $(1-\lambda)$ is there because of normalization. In the same way, the normalization factor in the SMA model is 1/N. In the SMA model, if I increase the window size N, then each observation carries a smaller weight: a smaller 1/N. Likewise, if I change $\lambda$ from 0.94 to 0.97 in EWMA, the effective window size increases (see Figure 3). As a result, each observation carries a smaller weight: a smaller $1-\lambda$.

- **SMA and EWMA:** The difference between these two volatility estimates becomes most visible immediately after a large price movement. Figure 4 uses the famous Black Wednesday of 1992 as an example to illustrate this point. This technical document by RiskMetrics was first written around 1994. If it were written today, then the 2008 crisis would be plotted here as an example.

  After a large price movement, up or down, the response of the EWMA estimate is very fast, because it carries a higher weight for the most recent event. If the market calms

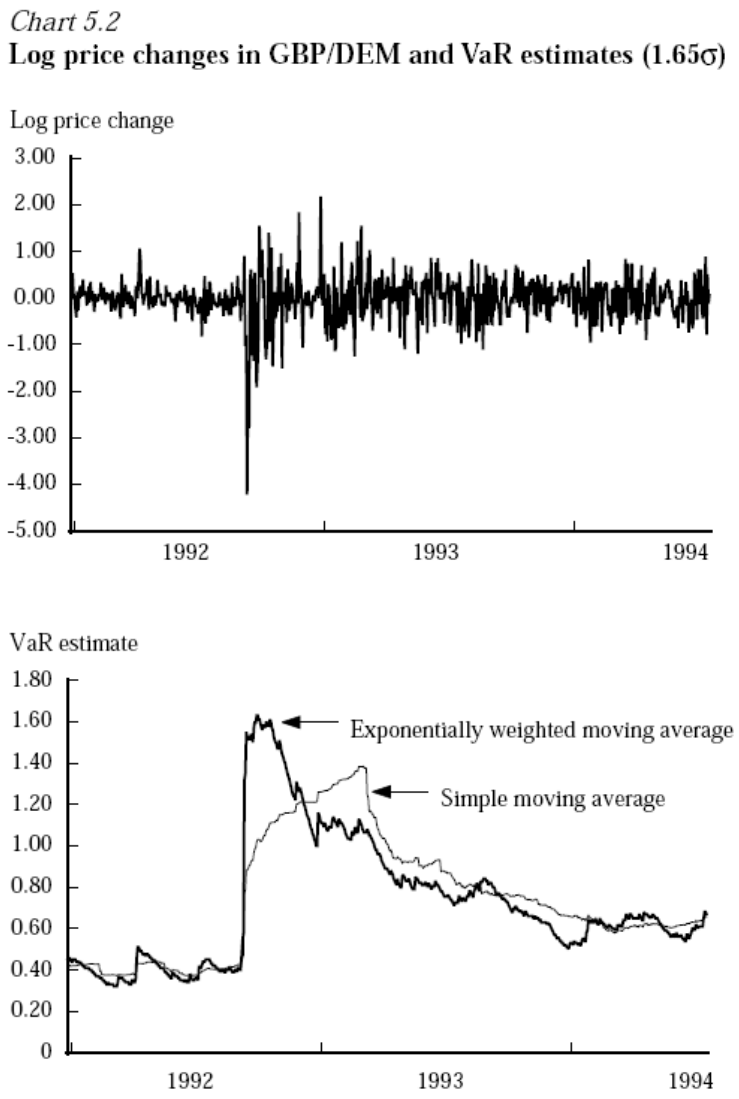


Figure 4: The Black Wednesday 1992 and the Volatility Estimates of SMA and EWMA.

down after the large price movement, then the EWMA estimate will soon come down to a lower level. The behavior of the SMA estimate is the opposite. Its response is typically sluggish and it carries that piece of information for the duration of its window size. For this reason, the EWMA is the preferred volatility estimate when it comes to monitoring market volatility at the daily frequency.

- **Black Wednesday 1992:** As a side, the 1992 sterling crisis was an important event in the global currency market. The followings are excepts from Steven Drobny's book on "Inside the House of Money: Top Hedge Fund Traders on Profiting in the Global Markets."

  The United Kingdom joined the European Exchange Rate Mechanism (ERM) in 1990 at a central parity rate of 2.95 deutsche marks to the pound. To comply with the ERM rule, the UK government was required to keep the pound in a trading band within 6 percent of the parity rate. In September 1992, as the sterling/mark exchange rate approached the lower end of the trading band, traders increasingly sold pounds against deutsche marks, forcing the Bank of England to intervene and buy an unlimited amount of pounds in accordance with ERM rules. Finally, on the evening of September 16, 1992, Great Britain humbly announced that it would no longer defend the trading band and withdrew the pound from the ERM system. The pound fell approximately 15 percent against the deutsche mark over the next few weeks, providing a windfall for speculators and a loss to the UK Treasury (i.e., British taxpayers) estimated to be in excess of £3 billion.

  It was reported at the time that Soros Fund Management made between $1-2 billion by shorting the pound, earning George Soros the moniker the man who broke the Bank of England. But he was certainly not alone in betting against the pound. In fact, the term global macro first entered the general public's vocabulary on Black Wednesday.

  Going back to our class on Predictability and Market Efficiency, there are few things to be learned. First, you predict the market by following the information, which, in this case, includes the ERM rule, the economic condition at UK, the government's ability and political resolve to defend its currency. Second, the "arbitrage" is risky. In order for George Soros to make $1 billion with a 15% drop in sterling, his short position at the time had to be over $6 billion. This is the style of global macro: large and risky directional bets. Of course, they don't always make money and we've seen a few times when Soros lost by the same order of magnitude. Third, most of the global macro opportunities (or losses) in currencies and emerging markets happened because of some

frictions outside of the financial markets: currency pegging, government intervention, central bank and policy errors, etc. As the governments and central banks become smarter in their interaction with the markets, such outsize returns may be slowly going away.

- **Computing EWA recursively:** Today is day $t-1$. Let $\sigma_t$ be the EWMA volatility estimate using all the information available on day $t-1$ for the purpose of forecasting the volatility on day $t$. Notice the dating convention: the time-$t$ estimate is observed on day $t-1$. In my personal opinion, we should date $\sigma_t$ by $t-1$, not $t$. But this is the convention in this area. So let's go with convention.

  Moving one day forward, it's now day $t$. After the day is over, we observe the realized return $R_t$. We now need to update our EWMA volatility estimator $\sigma_{t+1}$ using the newly arrived information (i.e. $R_t$):

$$\sigma_{t+1}^2 = \lambda\,\sigma_t^2 + (1-\lambda)\,R_t^2\,. \tag{1}$$

  A good exercise for you would be to start right from the beginning,

$$\sigma_2^2 = \lambda\,\sigma_1^2 + (1-\lambda)\,R_1^2$$

  and then apply the recursive formula a few times to convince yourself that this recursive approach does get you the exponential weighting scheme of EWMA:

$$\begin{aligned}
&\sigma_3^2 = \lambda\,\sigma_2^2 + (1-\lambda)\,R_2^2 = \lambda^2\,\sigma_1^2 + (1-\lambda)\left(\lambda R_1^2 + R_2^2\right)\\
&\sigma_4^2 = \lambda\,\sigma_3^2 + (1-\lambda)\,R_3^2 = \lambda^3\,\sigma_1^2 + (1-\lambda)\left(\lambda^2 R_1^2 + \lambda R_2^2 + R_3^2\right)\\
&\ldots\\
&\sigma_t^2 = \lambda^{t-1}\,\sigma_1^2 + (1-\lambda)\left(\lambda^{t-2} R_1^2 + \lambda^{t-3} R_2^2 + \ldots + R_{t-1}^2\right)
\end{aligned}$$

  For those of you who like things to be precise: as $t \to \infty$, we are back to the exact formulation of the EWMA. And whatever $\sigma_1$ we started with does not make a difference.

  If you are an Excel user, you will appreciate the convenience of this recursive formula. If you care about saving CPU time, you will also appreciate the convenience of this recursive formula. When we update the information on day $t$ to calculate $\sigma_{t+1}$, all of the past information has been neatly summarized by $\sigma_t$. The new information waiting for us to be included is the realization of $R_t$. We weight the new information $R_t^2$ by

$1-\lambda$ and "decay" the old information $\sigma_t^2$ by $\lambda$. Adding these two pieces together, we get the updated variance estimate. It would be difficult not to appreciate the elegance of this recursive approach. No?

- **The auto-correlation coefficient:** Another way to understand the recursive formula of Equation (1) is that imposes the dynamic structure of $\sigma^2$: persistent with an auto-correlation coefficient of $\lambda$.

  Recall that we regress stock return $R_{t+1}$ on its own lag $R_t$ to examine the stock return predictability. We find that from 1926 to 2004, the auto-correlation coefficient is positive and statistically significant. But the magnitude of the correlation is very small. Moreover, this predictability is not very robust: over the various subsamples, the auto-correlation coefficients become statistically insignificant. In other words, the random walk model with zero auto-correlation is a reasonable model for the stock returns.

  When it comes to the dynamic structure of volatility, however, the auto-correlation coefficient plays a rather important role. Models such as EWMA and GARCH became popular in practice because they allow volatility to be persistent with a high auto-correlation cofficient. In estimating the auto-correlation in stock returns, we can simply run a regression. In the case of volatility, however, we need to estiamte the volatility along with the coefficient $\lambda$. For this, we need a more structured estimation approach than a regression. (If you get into this area called Econometrics, you will realize that the essence is really the same. In particular, a linear regression is really the product of a maximum likelihood estimation. See Appendix A.)

- **Estimating the decay factor:** Figure 3 provides a graphical connection between the decay factor $\lambda$ and the sample size. A strong decay factor, say $\lambda = 0.8$, pays more attention to the current events and underweights the far-away events more strongly. As a result, the effective sample size is smaller with a stronger decay factor (e.g., smaller $\lambda$). As you can see from Figure 5, a strong decay factor improves on the timeliness of the volatility estimate, but the smaller sample size makes the estimate noisier and less precise. On the other hand, a weaker decay factor, say $\lambda = 0.97$, improves on the smoothness and precision, but that estimate could be sluggish and slow in response to changing market conditions, as reflected in Figure 5. So there is a tradeoff.

  - **Minimize RMSE:** Let's consider two ways to pick the optimal decay factor. In the first approach, we would like to minimize the forecast error between the model's prediction and the actual realization. Recall, on day $t$, we form $\sigma_{t+1}$ as a

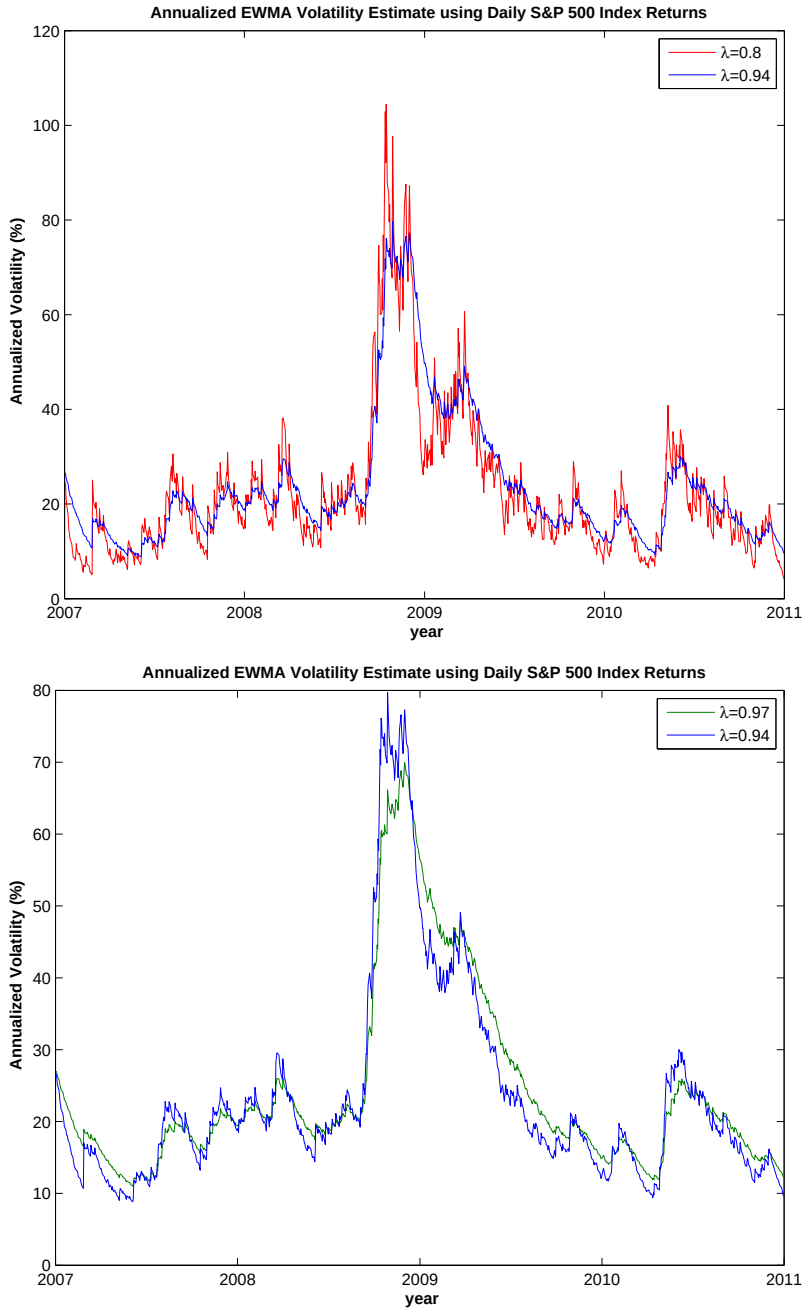


Figure 5: Time Series of EWMA Volatility Estimates with Varying Decay Factors.

forecast for the volatility on day $t+1$. So the model's forecast error is $R_{t+1}^2 - \sigma_{t+1}^2$. Summing these forecast errors over the sample period, we calculate the root mean squared error (RMSE) by,

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(R_{t+1}^2 - \sigma_{t+1}^2\right)^2}$$

Note that the only parameter at our disposal is $\lambda$. Everything else comes from the data. So let's find the optimal $\lambda^*$ that minimizes the forecast error:

$$\lambda^* = \arg\min_{\lambda\in(0,1)} \text{RMSE} = \arg\min_{\lambda\in(0,1)} \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(R_{t+1}^2 - \sigma_{t+1}^2\right)^2}$$

– **MLE:** In the second approach, let's use the maximum likelihood estimation. To be honest, using the MLE on this problem is really an overkill, but I would like to use this opportunity to introduce you to MLE. Anybody working with data should have done MLE at least once in their life.

Recall that we talk about the pdf of a normal, which is a Gaussian function. In our current setting, the volatility is time-varying. So the stock return $R_{t+1}$ is normally distributed only when conditioning on the volatility estimate $\sigma_{t+1}$:

$$f\left(R_{t+1}|\sigma_{t+1}\right) = \frac{1}{\sqrt{2\pi}\sigma_{t+1}} e^{-\frac{R_{t+1}^2}{2\sigma_{t+1}^2}}.$$

Notice that if I wanted to be very precise, I should have replaced $R_{t+1}^2$ by $(R_{t+1} - \mu)^2$ and use the MLE to estimate both $\lambda$ and $\mu$. But we talked about this. Setting $\mu = 0$ here is a good compromise to make.

The next step of MLE is to take log of the pdf:

$$\ln f\left(R_{t+1}|\sigma_{t+1}\right) = -\ln\sigma_{t+1} - \frac{R_{t+1}^2}{2\sigma_{t+1}^2},$$

I dropped $2\pi$ since it is a constant will not affect anything we will do later. We now add them up to get what econometricians call log-likelihood (llk):

$$\text{llk} = -\sum_{t=1}^{T}\left(\ln\sigma_{t+1} + \frac{R_{t+1}^2}{2\sigma_{t+1}^2}\right)$$

As you can see, the only parameter in llk is our choice of $\lambda$. It turns out that the best $\lambda$ is the one that maximizes llk. In practice, we take -llk and minimize -llk instead of maximizing llk.

What we just did came straight out of Econometrics. A good textbook on this topic is the Time Series Analysis by James Hamilton. Read in particular the chapter on Generalized Method of Moments. Most of the econometrics tasks we encounter in Finance can be understood from the perspective of GMM, which was developed by Prof. Lars Hansen at University of Chicago. Prof. Hansen shared the 2013 Nobel Prize with Prof. Eugene Fama and Prof. Robert Shiller. In the Appendix, I include my old PhD-era code for estimating the standard errors of mean, std, skewness, and kurtosis. As you can see, my approach was very much influenced by the GMM approach. In the Appendix, I wrote a brief note on MLE and linear regression, which could be a nice entry point to motivate you to learn more about Econometrics.

- **ARCH and GARCH:** The ARCH model, autoregressive conditional heteroskedasticity, was proposed by Professor Robert Engle in 1982. The GARCH model is a generalized version of ARCH. ARCH and GARCH are statistical models that capture the time-varying volatility:

$$\sigma_{t+1}^2 = a_0 + a_1 R_t^2 + a_2 \sigma_t^2$$

As you can see, it is very similar to the EWMA model. In fact, if we set $a_0 = 0$, $a_2 = \lambda$, and $a_1 = 1 - \lambda$, we are doing the EWMA model.

So what's the value added? This model has three parameters while the EWMA has only one. So it offers more flexibility (e.g., allows for mean reversion and better captures volatility clustering). If you are interested in estimating the GARCH model, you can use the MLE method we just discussed. Instead of estimating $\sigma_{t+1}$ using EWMA, you use the GARCH model. The EWMA has only one parameter $\lambda$ to estimate. The GARCH model has three parameters to estimate $a_0$, $a_1$, and $a_2$. You will find that, just like $\lambda$, $a_2$ is very close to one. In fact, $a_2$ captures the auto-correlation of the variance $\sigma_t^2$: an autocorrelation coefficient that is close to one indicates a very persistent time series. Moreover, after some calculation, you notice that the long-run mean of the variance in this model is $a_0/(1 - a_1 - a_2)$. You can see how having additional parameters could provide more flexibility to the model.

The GARCH model has a pretty strong influence, and you are encouraged to dig deeper

into the model if it interests you. We used to study quite a bit of GARCH at Stanford GSB. But looking back, I feel that I get most of the key intuitions by working with EWMA. Where there are too many moving parts and too many parameters, you tend to focus more on dealing with the formulas and parameters and lose track of the essence of the problem. That's why simplicity is always preferred.

# 3 EWMA for Covariance

As mentioned at the beginning, our goal is to create the variance-covariance matrix for the key risk factors influencing our portfolio. Suppose that there are two risk factors affecting our portfolios. Let $R_t^A$ and $R_t^B$ be the realized day-$t$ returns of these two risk factors. We estimate the covariance between A and B by

$$\text{cov}_{t+1} = \lambda \, \text{cov}_t + (1 - \lambda) \, R_t^A \times R_t^B$$

And their correlation:

$$\text{corr}_{t+1} = \frac{\text{cov}_{t+1}}{\sigma_{t+1}^A \sigma_{t+1}^B} \, ,$$

where $\sigma_{t+1}^A$ and $\sigma_{t+1}^B$ are the EWMA volatility estimates.

This calculation of covariance and correlation is pretty straightforward once you master the EWMA recursive formula. But let me use this opportunity to bring in volatility as a risk factor and emphasize on its importance. As recent as the early 2000s, volatility as a risk factor was not widely monitored by market participants. Of course, sophisticated investors pay attention to their exposure to volatility risk. The general intuition is that if you are short on volatility, you are going to lose during crisis. On the other hand, if you are long on volatility, you are partially hedged during these crises. Exposures to volatility risk comes certain non-linearity in one's position. The most straightforward way to be long on volatility is to buy at-the-money S&P 500 index options. As we will cover in our options class, such long positions usually are expensive. That is, you are paying a premium for such positive exposures.

Since the 2008 financial crisis, the volatility risk has got a broader audience. By now, the VIX index is reported daily in a prominent position along with the Dow, the S&P, and Nasdaq. It's often called the fear gauge. Figure 6 plots the historical VIX for the past 15 years. Going over the various events in the past, you can certain appreciate why it is called fear gauge.

Another important observation about the volatility risk factor is its increasing negative

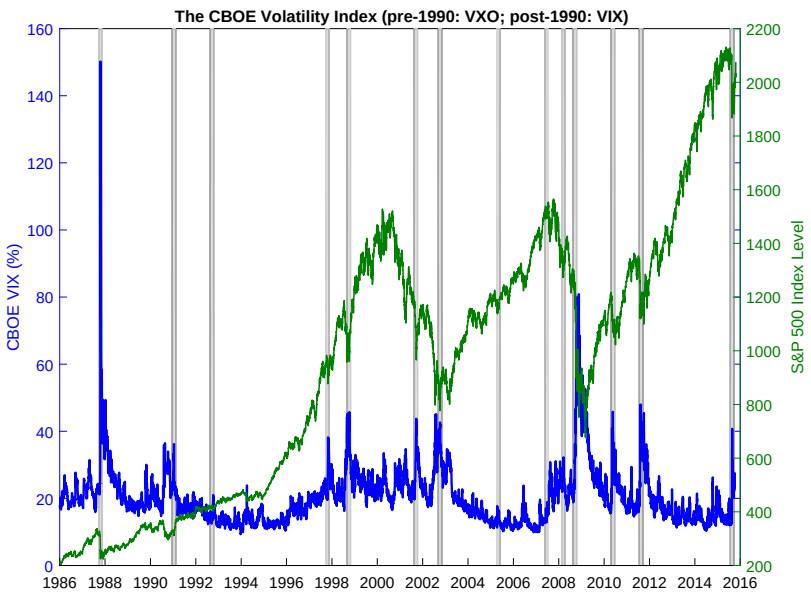


Figure 6: Time Series of CBOE VIX.

correlation between the market risk. Figure 7 plots the EWMA correlation estimate between daily returns of the S&P 500 index and the daily changes in VIX. As you can this, this correlation is always negative. The fact that there is a negative correlation between the stock market return and its volatility has always been documented, well before CBOE published its VIX index. As you can see, during the early sample period, the correlation was hovering around -50%. Was I was a PhD student working on this topic in the late 1990s, a typical number for this correlation would be -60%. There is certainly a trend of this correlation becoming more negative in recent times. After the 2008 financial crisis, this correlation certainly has experienced a regime switch to a more negative territory.

A negative correlation implies that whenever market drops down, the volatility goes up. Using the interpretation of VIX as a fear gauge, this means that a down market is coupled with increasing fear. The more negative correlation in recent years means a higher level of sensitivity to down markets: a market sitting at its edge, more easily spooked. As we move on to the options market, we will look at the "fear" component in VIX more closely.

# 4 Calculating Volatility and VaR for a Portfolio

- **Portfolio volatility:** Suppose that our portfolio has only two risk factors, whose daily returns are $R^A$ and $R^B$, respectively. Suppose we've done our risk mapping from individual positions to portfolio weights on these two risk factors: $w_A$ and $w_B$. If we

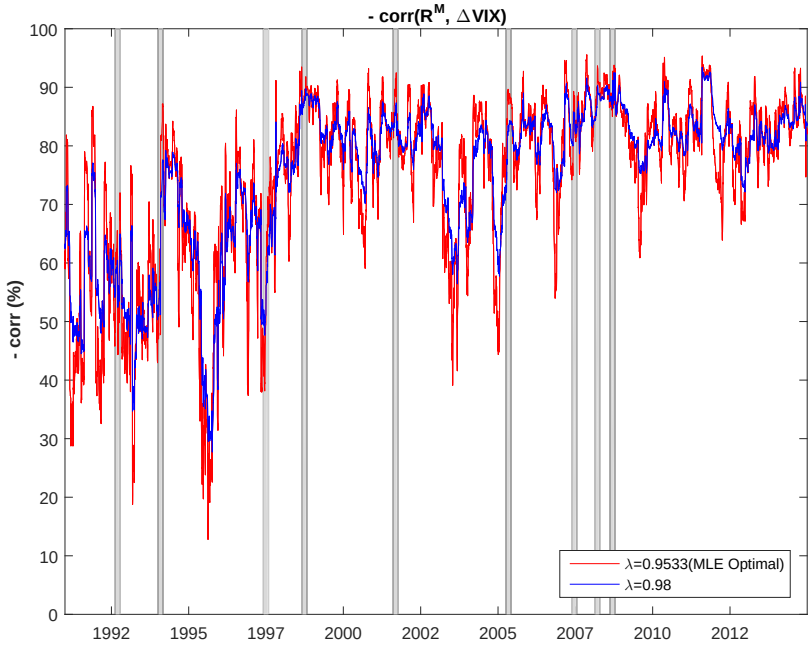


Figure 7: Time Series of CBOE VIX.

focus our attention on the risk part of our portfolio, then we can even normalize the portfolio weights so that $w_A + w_B = 1$. Let's also assume that at the moment, our risk portfolio has a market value of \$100 million.

We first construct a variance-covariance matrix for our risk factors:

$$\Sigma_t = \begin{pmatrix} (\sigma_t^A)^2 & \rho_t^{AB}\sigma_t^A\sigma_t^B \\ \rho_t^{AB}\sigma_t^A\sigma_t^B & \left(\sigma_t^B\right)^2 \end{pmatrix}$$

It is a 2×2 matrix, since we have only two risk factors. If you have 100 risk factors in your portfolio, then you will have a 100×100 matrix. For example, in JPMorgan's RiskMetrics, roughly 480 risk factors were used. In Goldman's annual report, 70,000 risk factors were mentioned. A risk manager deals with this type of matrices everyday and the dimension of the matrix can easily be more than 100, given the institution's portfolio holdings and risk exposures. Notice also the timing here. For $\sigma_t$, you are actually using all of the market information on day $t-1$ (e.g., daily returns of assets A and B up to day $t-1$), for the purpose of forecasting volatility for day $t$.

Let's time-stamp our portfolio weights by the actual time. Suppose today is $t-1$ and

let the portfolio weight be written in a vector form:

$$w_{t-1} = \begin{pmatrix} w_{t-1}^A \\ w_{t-1}^B \end{pmatrix}$$

Our portfolio volatility is

$$\sigma_t^2 = \begin{pmatrix} w_{t-1}^A & w_{t-1}^B \end{pmatrix} \times \begin{pmatrix} (\sigma_t^A)^2 & \rho_t^{AB}\sigma_t^A\sigma_t^B \\ \rho_t^{AB}\sigma_t^A\sigma_t^B & \left(\sigma_t^B\right)^2 \end{pmatrix} \times \begin{pmatrix} w_{t-1}^A \\ w_{t-1}^B \end{pmatrix}$$

Using the notation we've developed so far, we can also write

$$\sigma_t^2 = w_{t-1}' \times \Sigma_t \times w_{t-1}\,,$$

which involves using mmult and transpose in Excel.

- **Portfolio VaR:** Suppose that our daily portfolio volatility is $\sigma$ (daily number, unannualized). The value of our portfolio, marked to the market, is \$100 million. Assuming that the portfolio return is normally distributed, we can estimate how much we stand to lose in market value if a 5% tail event happens to our portfolio over the next day:

  $$\text{VaR } (95\%) = \text{portfolio value} \times 1.645 \times \sigma\,,$$

  where 1.645 is the critical value for a 5% tail event. Some firms report 99% VaR, which corresponds to the loss in market value if a 1% tail event happens to the portfolio over the next day: our portfolio over the next day:

  $$\text{VaR } (99\%) = \text{portfolio value} \times 2.326 \times \sigma\,,$$

  where 2.326 is the critical value for a 1% tail event.

  As you can see, there are two main drivers for the portfolio VaR: the market value of the portfolio and the portfolio volatility. The market value tells you the dollar exposure of your firm's trading book to risky assets and the portfolio volatility tells you how volatile the risky assets are. For a chief executive of a firm, the VaR number is a useful summary of these two important components of a firm's trading book. Although VaR is framed as a consideration over tail events, it is not really a measure of tail risk since it is driven by volatility. We will come back to this issue again when we spend a class on Market Risk Management.

# Appendix

## A MLE and Linear Regression

Let's consider the linear regression:

$$Y_t = \alpha + \beta X_t + \epsilon_t \,,$$

where if we replace X with $R^M - r_f$ and Y with $R^i - r_f$, we are back with our favorite CAPM regression.

Thinking in terms of MLE, we focus on the distribution of the regression residual $\epsilon_t$. We assume that $\epsilon_t$ is i.i.d. with normal distribution: zero mean and volatility of $\sigma_\epsilon$. Now let's repeat the MLE steps for this regression:

- We write down the pdf for the residual:

$$f(\epsilon_t) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} e^{-\frac{\epsilon_t^2}{2\sigma_\epsilon^2}}$$

- Take the log of the pdf:

$$\ln f(\epsilon_t) = -\ln \sigma_\epsilon - \frac{\epsilon_t^2}{2\sigma_\epsilon^2} = -\ln \sigma_\epsilon - \frac{(Y_t - \alpha - \beta\, X_t)^2}{2\sigma_\epsilon^2} \,,$$

  where $2\pi$ was again dropped.

- Summing up all observations to get

$$\text{llk} = \sum_{t=1}^{T} \left( -\ln \sigma_\epsilon - \frac{(Y_t - \alpha - \beta\, X_t)^2}{2\sigma_\epsilon^2} \right)$$

- Find the parameter values ($\sigma_\epsilon$, $\alpha$, and $\beta$) that will minimize this,

$$-\text{llk} = T \times \ln \sigma_\epsilon + \frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^{T} (Y_t - \alpha - \beta\, X_t)^2$$

In the EWMA case, we use the computer to minimize -llk by varying $\lambda$. In the GARCH case, we use the computer to minimize -llk by varying $a_0$, $a_1$ and $a_2$. Here, we can actually

do it by hand. Let's forget $\sigma_\epsilon$ for now, and focus on $\alpha$ and $\beta$. To miminize $-$llk is the same as finding $\alpha$ and $\beta$ so that

$$\frac{\partial \,\text{llk}}{\partial \alpha} = \frac{\sum (Y_t - \alpha - \beta\, X_t)}{\sigma_\epsilon^2} = 0$$

and

$$\frac{\partial \,\text{llk}}{\partial \beta} = \frac{\sum \left[X_t\ (Y_t - \alpha - \beta\, X_t)\right]}{\sigma_\epsilon^2} = 0\,.$$

Solving for the optimal $\alpha$ and $\beta$ then reduces to solving the above two equations. The first derivatives of llk with respect to the model parameters (e.g., $\alpha$ and $\beta$) are also called score. If there are two parameters, then the score is a vector of two. In our current case, the score should be a vector of three because there are three parameters: $\alpha$, $\beta$, and $\sigma_\epsilon$. But as agreed, let's focus only on $\alpha$ and $\beta$ and forget about $\sigma_\epsilon$.

Solving for the partial derivative (score) with respect to $\alpha$, we have,

$$\alpha = \frac{1}{T}\sum Y_t - \beta\,\frac{1}{T}\sum X_t$$

Solving for the partial derivative with respect to $\beta$, we have

$$\sum X_t\, Y_t - \alpha \sum X_t - \beta \sum X_t^2 = 0$$

Plugging the solution for $\alpha$ into the equation above, we have:

$$\sum X_t Y_t - \frac{1}{T}\sum X_t \sum Y_t = \beta\left(\sum X_t^2 - \frac{1}{T}\left(\sum X_t\right)^2\right)$$

Let me divide both sides of the equation by $T$ so that you can see the result more clearly,

$$\frac{1}{T}\sum X_t Y_t - \left(\frac{1}{T}\sum X_t\right)\left(\frac{1}{T}\sum Y_t\right) = \beta\left(\frac{1}{T}\sum X_t^2 - \left(\frac{1}{T}\sum X_t\right)^2\right)$$

What we have is,

$$\text{cov}(X, Y) = \beta\,\text{var}(X)$$

So, as a by product of our derivation, you get to know why running a regression gets you the CAPM beta.

Also, for those of you who think more carefully, the fact that we assume $\epsilon$ is normally distrubted might be bothering you. Don't be. Even if $\epsilon$ is not normally distributed, we can

still do this procedue, which is called quasi-maximum likelihood estimation. The estimates might not be the most efficient, but they are still consistent. By going through this derivation, my intension is to lead some of you to the door of econometrics. If you are interested, go ahead. If not, turn around. One key calculation left out is how to calculate the standard errors of $\alpha$ and $\beta$. For those of you who are interested in learning more, I would recommend the chapters on GMM and MLE of James Hamilton's book on Time Series Analysis.

# B Matlab Code

Code 1: Plot SMA Volatility Estimates

```
load SP500_Daily.txt;
Data=SP500_Daily;

yr=Data(:,1);
mn=Data(:,2);
dy=Data(:,3);
Time=datenum(yr,mn,dy);

Ret=Data(:,4)*100;

time_mn=[]; vol_mn=[]; mu_mn=[];
for i_yr=min(yr):1:max(yr),
   for i_mn=1:12,
       i_Ret=Ret(yr==i_yr&mn==i_mn);
       if ~isempty(i_Ret),
          time_mn=[time_mn; datenum(i_yr,i_mn,30)];
          [M,SE]=stat_fun(i_Ret);
          vol_mn=[vol_mn; [std(i_Ret)*sqrt(252) SE(2)*sqrt(252)]]; %monthly
              vol estimate with standard error
          mu_mn=[mu_mn; [mean(i_Ret) std(i_Ret)/sqrt(length(i_Ret))]]; %
              monthly mu estimate with standard error
       end
   end
end

% plot SMA vol estimates
figure(2);
```

```
plot(time_mn,vol_mn(:,1),'b-');
hold on;
datetick('x','yyyy')
BND=axis;
axis([datenum(1962,1,1) datenum(2011,12,31) BND(3) BND(4)]);
BND=axis;
plot([BND(1) BND(2)],std(Ret)*sqrt(252)*[1 1],'k--')
hold off

% plot SMA vol estimates with confidence intervals
figure(10);
plot(time_mn,vol_mn(:,1)+1.96*vol_mn(:,2),'g-',time_mn,vol_mn(:,1)-1.96*
    vol_mn(:,2),'m-');
hold on
plot(time_mn,vol_mn(:,1),'b-','LineWidth',2);
hold off
BND=axis;
datetick('x','yyyy')
legend('95% Confidence, Upper','95% Confidence, Lower')
title('\bf SMA estimates of \sigma and their 95% confidence intervals');
ylabel('\bf Annualized Volatility (%)');
BND=axis;
axis([datenum(1962,1,1) datenum(2011,12,31) BND(3) BND(4)]);

% plot SMA mu estimates with confidence intervals
figure(11);
plot(time_mn,mu_mn(:,1)+1.96*mu_mn(:,2),'g-',time_mn,mu_mn(:,1)-1.96*mu_mn
    (:,2),'m-');
hold on
plot(time_mn,mu_mn(:,1),'b-','LineWidth',2);
hold off
BND=axis;
datetick('x','yyyy')
legend('95% Confidence, Upper','95% Confidence, Lower')
BND=axis;
axis([datenum(1962,1,1) datenum(2011,12,31) BND(3) BND(4)]);
ylabel('\bf Monthly Average of Daily Returns (%)');
title('\bf SMA estimates of \mu and their 95% confidence intervals');
```

```
% plot SMA vol estimates together with NBER recessions
figure(3);
plot(time_mn,vol_mn(:,1),'b-');
BND=axis;
hold on;
datetick('x','yyyy')
FY=[BND(4) BND(3) BND(3) BND(4)];
load NBER_Recession.dat;
hold on
for i=1:size(NBER_Recession,1),
  FX=[datenum([NBER_Recession(i,1:2) 1])*[1 1] ...
      datenum([NBER_Recession(i,3:4) 1])*[1 1]];
  if FX(1)> datenum(1962,1,1),
    fill(FX,FY,[0.75 0.75 0.75]);
    hold on
  end
end
plot(time_mn,vol_mn(:,1),'b-','LineWidth',2);
hold on;
plot([BND(1) BND(2)],std(Ret)*sqrt(252)*[1 1],'k--')
hold off;
BND=axis;
axis([datenum(1962,1,1) datenum(2011,12,31) BND(3) BND(4)]);
```

Code 2: Calculating Standard Errors

```
function [MOMENTS, SE]=my_stat(data)

T=size(data,1);

m1=mean(data);
m2=var(data);
m3=mean((data-m1).^3);
m4=mean((data-m1).^4);

MEAN=m1;
STD=sqrt(m2);
```

```
SKEW=m3/m2^(3/2);
KURT=m4/m2^2;


h1=data-m1;
h2=h1.^2-m2;
h3=h1.^3-m3;
h4=h1.^4-m4;
h=[h1 h2 h3 h4];
T=length(h);
R=h'*h/T;
n_moving=5;
for i=1:n_moving
   R_temp=h(i+1:T,:)'*h(1:T-i,:)/T;
   R=R+(R_temp'+R_temp)*(1-i/(n_moving+1));
end
W=inv(R);
D=[-1 0 0 0; 0 -1 0 0; 3*m2 0 -1 0; 4*m3 0 0 -1];

COV=inv(D'*W*D);
SE=sqrt(diag(COV)/T);

D2=1/2/sqrt(m2);
C2=COV(2,2);
SE(2)=sqrt(D2*C2*D2/T);
D23=[-1.5*m3/m2^(5/2) 1/m2^(3/2)];
C23=COV(2:3,2:3);
SE(3)=sqrt(D23*C23*D23'/T);
D24=[-2*m4/m2^3 1/m2^2];
C24=COV([2 4],[2 4]);
SE(4)=sqrt(D24*C24*D24'/T);

MOMENTS=[MEAN STD SKEW KURT]';
```

Jun Pan
junpan@mit.edu

253-3083
E62-624

## Class 8: Equity in the Time Series, Part 1

## Predicting the Market

This Version: September 27, 2016

For some, predicting the market is a safe conversation piece, just like talking about the weather. I became a Finance professor in July 2000. The day before, I was a PhD student. The day after, I became a professor. Nothing really changed. But because of the new label, suddenly people were seeking for my opinion on financial matters. By far, the question I got the most was can I teach them how to predict the market. Wanting to know something about the future is hard wired in most of us. On Wall Street, the appetite for predicting stock prices is as old as the existence of the markets. In this class, let's take a look at the empirical evidences on stock return predictability.

We will start with the efficient market hypothesis, using it as a framework to help us understand what it means to be able to predict the market. People often believe that market efficiency means that returns are unpredictable. This is not true. In an economy with time-varying business condition or time-varying risk appetite, the expected returns are time-varying and, most likely, persistent. As a result, you will see return predictability. The more relevant question is: How strong is the predictability? We will look at some of the empirical evidences.

# 1 Predictability and Market Efficiency

- **Follow the information:** The financial markets are an information central, where people bring their information to trade. If it is a correct and useful piece of information, which has not yet been incorporated into the price, then there is room for profit. But as soon as the market price adjusts to the news, the information loses its usefulness and there is no longer any profit to be made with this piece of information.

  So when it comes to predicting the market, one should follow the flow of information. A trader who wants to make a profit from predicting the market should always ask himself: Am I good at collecting information? If so, then I have all the incentive to do

so because I will be rewarded for bringing this information to the financial markets. The next question is: What kind of information am I good at collecting, macro-level for the entire economy, or micro-level for individual stocks? Depending on your talent, the nature of your trading strategy will be very different. Global macro funds place directional bets on the overall market: interest rate, foreign exchange, and maybe the stock market. Long/short equity funds or fixed-income arbitrage funds avoid taking any directional bet. Instead, they focus on the relative mis-pricing between groups of stocks or bonds. At the super high-frequency domain, where the life span of information is on the order of milliseconds, market making funds and statistical arbitrage funds populate this space to facilitate trades and provide liquidity.

All of these market players are motivated by a common goal: making a profit. And they are able to do so by bringing information to the markets. As a result of these efforts, new and relevant information gets incorporated into the prices. And the markets become more efficient.

- **The efficient market hypothesis:** It's impossible to talk about market predictability without bringing up the efficient market hypothesis, or the question about market efficiency. So let me spend some time clarifying some of the confusions.

  First of all, in my personal opinion, the efficient market hypothesis is simply a statement that defines what it means to have an efficient market: when market prices incorporate information. It is like saying, being happy means to have peace of mind.

  Second, without a proper asset pricing model, there is no way to test the efficient market hypothesis. With a proper model, then we are simply testing the model (e.g., the CAPM) which usually assumes market efficiency. So there is really no point in sweating over it. To be more specific, the efficient market hypothesis is not a stand alone test on market efficiency. It is always a joint test. Market efficiency can only be tested in the context of an asset pricing model that specifies equilibrium expected returns. For example, market efficiency implies zero predictability only if the expected returns that investors require to hold stocks are constant through time (or at least serially uncorrelated). Otherwise, if expected stock returns are time-varying and persistent, then there will be predictability in stock returns and it does not imply at all market inefficiency.

  Third, Finance in general and efficient market hypothesis in particular is really not a system of beliefs. What we can offer in Finance are tools. Tools for clear thinking. Don't believe, don't don't believe. Use the tools, apply them to the data and to your

own experiences, make an honest and sincere effort to figure things out for yourself.

- **Orange juice:** Since we are on the topic of market efficiency, let me tell one story that impressed me the most over the years. It is about orange juice, written in a 1984 paper by Prof. Roll from UCLA. It is the kind of paper I've always wanted to write: simplicity at its best; maximum power with minimum fluff.

  Cold weather is bad for orange production. Orange trees cannot withstand freezing temperatures that last for more than a few hours. The central Florida region around Orlando, which accounts for more than 98 percent of U.S. production of frozen concentrated orange juice, occasionally has freezing weather. During the 6 and 1/4 year period studied by Prof. Roll, there were four periods when the temperatures were below 30°F, each accompanied by significant price increase in orange juice futures prices.

  Overall, the most important determinant in the pricing of orange juice futures is weather in central Florida. Quoting Prof. Roll, "So if the OJ futures market is an efficient information processor, it should incorporate all publicly available long-term and short-term weather forecasts. Any private forecasts should be incorporated to the extent that traders who are aware of those forecasts are also in command of significant resources. The futures price should, therefore, incorporate the predictable part of weather in advance."

  With this idea in mind, Prof. Roll uses the OJ futures prices to predict the weather. Not surprisingly, you will find a relationship between the two. The ingenious design of Prof. Roll's regression is to find out if the OJ futures prices can predict weather more accurately than the National Weather Service. On the left hand side of his regression is the temperature forecast error, which the percentage difference between the actual temperature and the forecast temperature provided by the National Weather Service. On the right hand side of his regression is the returns on orange juice futures.

  What did he find? Orange juice futures prices are better at predicting the weather than the National Weather Service. This predictability is especially strong for the P.M. tempature forecast because of the sensitivity of orange trees to freezing temperatures.

- **The value of millisecond:** Let me tell you another story that fascinated me. It borderlines on craziness, but is a good story. This the first paragraph of *Flash Boys*, a recent book by Michael Lewis.

  "By the summer of 2009 the line had a life of its own, and two thousand men were

digging and boring the strange home it needed to survive. Two hundred and five crews of eight men each, plus assorted advisors and inspectors, were now rising early to figure out how to blast a hole through some innocent mountain, or tunnel under some riverbed, or dig a trench beside a country road that lacked a roadside – all without ever answering the obvious question: *Why?* The line was just a one-and-a-half-inch-wide hard black plastic tube designed to shelter four hundred hair-thin strands of glass, but it already had the feeling of a living creature, a subterranean reptile, with its peculiar needs and wants. It needed its burrow to be straight, maybe the most insistently straight path ever dug into the earth. It needed to connect a data center on the South Side of Chicago to a stock exchange in northern New Jersey. Above all, apparently, it needed to be a secret."

All of these effort just so the speed of information transmission can be improved in the order of ... millisecond. Let me quote Lewis again, since he is a much better writer.

"One way to price access to the line, Tabb thought, was to figure out how much money might be made from it, from the so-called spread trade between New York and Chicago – the simple arbitrage between cash and futures. Tabb estimated that if a single Wall Street bank were to exploit the countless minuscule discrepancies in price between Thing A in Chicago and Thing A in New York, they'd make profits of $20 billion a year. He further estimated that there were as many as four hundred firms then vying to capture the $20 billion."

- **Market efficiency is not a marble statue:** In telling the previous two stories, I would like to impress upon you the process through which markets become efficient. Market efficiency is not really a doctrine for you to believe or disbelieve. It is a process, a process of arbitrageurs participating in the markets with the objective of making a profit. Sometimes, this process works; sometimes it fails. It is an organic process, not a marble statue.

  After the 2008 financial crisis, many people were hard on the efficiency market hypothesis. Some people believed that the financial crisis was the result of a misguided faith in market efficiency that encouraged market participants to accept security prices as the best estimate of value rather than conduct their own investigation. Some wrote that among the causes of the recent financial crisis was an unjustified faith in rational expectations, market efficiencies, and the techniques of modern finance.

  Seriously, I really don't know how these people got their ideas. Rational expectation builds on the understanding that all players in the market are motivated to optimize

their risk and return tradeoff; market efficiency does not happen in the vacuum; it happens only when investors bring their information to the market with the objective of making a profit; and techniques of modern finance do help reduce trading cost and improve risk sharing in the society.

As to 2008? The flow of information broke down at some point. Large banks were sitting on supposedly super safe tranches of CDO and CDO2 without realizing or the willingness to realize the real risk. The rest of the market had a very limited access to this kind of balance-sheet (or off balance-sheet) level information and the market prices failed to incorporate this information. But did the banks take these positions out of their belief of market efficiency? I really doubt it.

- **Market inefficiency and limits to arbitrage:** Since we touched upon the topic of market efficiency, I think it would be fair to mention the Behavior Finance literature on market inefficiency. It was an area of Finance that grew in popularity after the tech boom of 1990s. If you are interested in this topic, you can start with Prof. Shleifer's book, "Inefficient Markets: An Introduction to Behavioral Finance."

  The efficient market hypothesis assumes that the market incorporates the new information right away. In practice, however, there is uncertainty surrounding the information and the process of price discovery itself involves uncertainty. In certain situations, a correct piece of information might not get incorporated into the price right away. If the price moves temporarily in the opposite direction of his information, the arbitrageur might in fact lose money trading this information. This argument, proposed by Shleifer and Vishny (1997) and often referred to as "limits to arbitrage" can help explain why bubble can keep building up even when many people are calling it a bubble.

  When Alan Greenspan, the then chairman of the Fed, gave the famous "irrational exuberance" speech in December 1996, the Nasdaq was around 1,300. Initially, the stock markets around the world dropped precipitously in reaction to the speech. But the markets soon shrugged off the warning and started the most spectacular upward trajectory in the history of Nasdaq. A little over three years after the speech, on March 10, 2000, the Nasdaq peaked at 5,048.62. Then it went down as fast as it came up, and bottomed near 1,140 two and half years later on October 4, 2002.

  One person who shared the same view with chairman Greenspan was Prof Shiller, who later wrote a book titled "irrational exuberance." Prof Shiller also shared the 2013 Nobel Prize with Prof Fama and Hansen. It was said that Prof Shiller, following his own prediction about the internet bubble, actually shorted the Nasdaq in the late

1990s, only to lose money because the market kept its upward trajectory for too long and crashed much too late.

# 2 Predicting the Market

- **What we've learned so far?** Talking about market efficiency and market predictability at a hypothetical level is just not that interesting. Now that we are seven classes into the semester, maybe we can start from what we've learned so far.

  By learning about the various quant strategies, we do recognize that the alpha generated by a quant strategy does come from a certain ability to predict the future. People might vary in their opinion on whether the alpha comes from market inefficiency (under/over-reactions) or systematic risk exposure. One observation I am sure that you've made is that quant investors do not take a stand on the market risk. If possible, they choose to avoid the market risk by taking long/short positions of two portfolios with similar beta exposures.

  And yet, the market risk remains the most important and pervasive. You've probably noticed in our Assignment 1 that market-neutral hedge funds are not really market neutral. For example, the hedge fund index in long/short equity has a beta around 0.40. Even for market-neutral hedge funds, the beta exposure is non-zero: around 0.20. Overall, the market risk is an important risk and let's try to understand it more.

  In this class, we will focus on the "first moment" of the aggregate stock market and move on to the "second moment" in the next class.

- **How good are investors at predicting the market?** You must have heard this famous story about Rockefeller and his shoe shine boy. After receiving unsolicited stock tips from his shoe shine boy in 1928, Rockefeller decided to get out of the stock market. His rational: when a shoe shine boy started to give stock tips, the market probably was reaching its peak.

  I don't know if the story actually happened to Rockefeller, but the gist of the story got repeated again and again in the history of financial markets. Last year, from July 2014 to July 2015, I was on sabbatical and spent most of my time in Shanghai with my parents. I was living a very simple life, far away from the financial establishments in Shanghai. Yet one can hardly avoid the hype and then the disappointment of the stock market. I had to tell my 80-year old father repeatedly that his optimal allocation

to the stock market is zero, regardless of how much money other people were making out of the market.

The empirical evidence paints a similar story: investors have no ability to predict the future. In fact, their prediction is a response to the stock market. When the markets are doing well, their prediction is optimistic; when the markets are doing poorly, they become pessimistic. Moreover, their prediction affects their behavior. The flow to equity mutual fund is driven heavily by the recent stock market performance. The same pattern of flow chasing performance can also be found in bond mutual funds.

- **Use past returns to predict the future:** For anyone wanting to predict the stock market, probably the very first regression would be:

$$R_{t+1} = a + \rho R_t + \epsilon_{t+1} \, .$$

Given the time-series data, this is the easiest regression to run. The results are mixed, depending on the horizon over which this regression is run. At the monthly horizon, the autocorrelation $\rho$ is generally positive and statistically significant. The magnitude is small for the value-weighted portfolio and becomes larger for the equal weighted portfolio. But this result is not very stable and could flip sign or become insignificant during sub-sample analyses. Overall, the R-squared of this predictive regression is very small, indicating that much of the future returns remains unpredictable.

In academic, there is a pretty large literature on this topic. If you are interested, you can read "Permanent and Temporary Components of Stock Prices" by Fama and French (1988), who ran this regression over a horizon of 3-5 years and found large negative autocorrelations. In "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," Lo and MacKinlay (1988) propose the innovative variance ratio test as an alternative to the regression analysis. In "When are Contrarian Profits due to Stock Market Overreaction?" Lo and MacKinlay (1990) explain that despite negative autocorrelation in individual stock returns, weekly portfolio returns are positively autocorrelated and are the result of important cross-autocorrelations.

- **Stock returns and business cycle:** In any Finance model, one main driver for stock returns should be the underlying economic condition. Nevertheless, the link between the two is not that strong in the data. For example, Prof. Shiller wrote a paper in 1984 entitled, "Do Stock Prices Move Too Much to be Justified by Subsequent Change in Dividend?" In this paper, he made the observation that the stock market is too

volatility (e.g., 20% per year) compared to the volatility in the fundamental: dividends or earnings.

If you plot the time-series of realized stock returns against the business cycle, you do find a link between the two. In particular, depressed expected business conditions are associated with high expected excess returns. This observation gives rise to predictive regressions using a set of variables that are related to business conditions, including default spreads, term premiums, and dividend-price ratio. By far, the best predictor for stock market returns is the dividend-price ratio. We will re-visit the default spread and term premium as we cover the fixed-income market.

- **Dividend-price ratio as a stock market predictor:** Let's run this regression at the annual frequency:

$$R_{t+1} = a + b\left(\frac{D}{P}\right)_t + \epsilon_{t+1}\,,$$

where $D/P$ is the dividend-price ratio (aggregate dividend divided by the value-weighted CRSP index). The general finding that is the coefficient $b$ of this predictive regression is positive and statistically significant.

The key to this regression is 1/P. The aggregate price level is usually depressed during poor business condition (e.g. recessions). Going forward, the stock return is expected to be high. Hence the positive regression coefficient. Using D/P is just a way to scale the overall time trend of stock price increase. Using aggregate earnings, one can use replace D/P by E/P, although the earnings number is more noisy and biases the regression coefficient downward.

One important observation of this predictive regression is that the power of predictability is very weak even for the best predictor. At an annual frequency, the R-squared of this predictive regression is around 5%, indicating that 95% of the future variance remains unpredictable.

- **Market timing as a trading strategy:** This kind of result explains why market timing is not a very popular trading strategy among market participants. Of course, there is nothing wrong with taking a long position on the market if your objective is to be compensated from such a risk exposure. But if you are in and out of the market with the belief that you can predict the market, then you should be very careful. The empirical evidence tells you that you are going to be exposed to quite a bit of uncertainty. Moreover, this is a very special kind of uncertainty — the market risk, the most dangerous kind.

This is why David Swensen wrote in his book, "Market timing, according to Charles Ellis, represents a losing strategy: There is no evidence of any large institutions having anything like consistent ability to get in when the market is low and get out when the market is high. Attempts to switch between stocks and bonds, or between stocks and cash, in anticipation of market moves have been unsuccessful much more often than they have been successful. Serious investors avoid timing markets."

Jun Pan
junpan@mit.edu

253-3083
E62-624

## Classes 12 & 13: Options, Part 1

This Version: October 19, 2016

# 1 Options, an Overview

- **Why Options?** The development of options as an exchange-traded product was an important landmark in the practice of Finance. It offers investors an alternate way to buy and sell the risk inherent in the underlying stock. In the language developed later, it offers non-linear exposures to the underlying stock or index. This non-linearity cuts the entire distribution of stock returns into various pieces.

  After the 2008 crisis, people sneered at the Wall Street practices such as tranching and repackaging. I think this is very unfortunate. A small fraction of Wall Street clearly mis-used and abused derivatives and contributed to the financial crisis in 2008. Looking back into the history of financial innovation, this was not the first time, nor will it be the last.

  Finance is about optimal allocation of risk: match the right kind of risk to the right kind of investors and distribute the right kind of investment to the right kind of firms or entrepreneurs. If we think of this distributional effort as a network of pipelines, then financial markets on equity, bond, foreign exchange, and commodity offer the basic infrastructure. The limited flexibility of these markets gave rise to derivatives.

  Options are a very good example. When we invest in the stock market, we have to take the whole package: the entire distribution of the stock. In our earlier classes, we talked about how we can minimize our exposure to idiosyncratic risk by forming portfolios and how we can take out the market risk by long/short strategies. The motivation behind options is the same. What if we are interested in hedging out not the entire market risk, but only a specific portion of the market risk, say the left tail? The long/short strategy will not help us do that: you are either all in or all out. But buying a put option on the S&P 500 index will achieve this goal for you. Now the question is how much are you willing to pay for this product? This is option pricing.

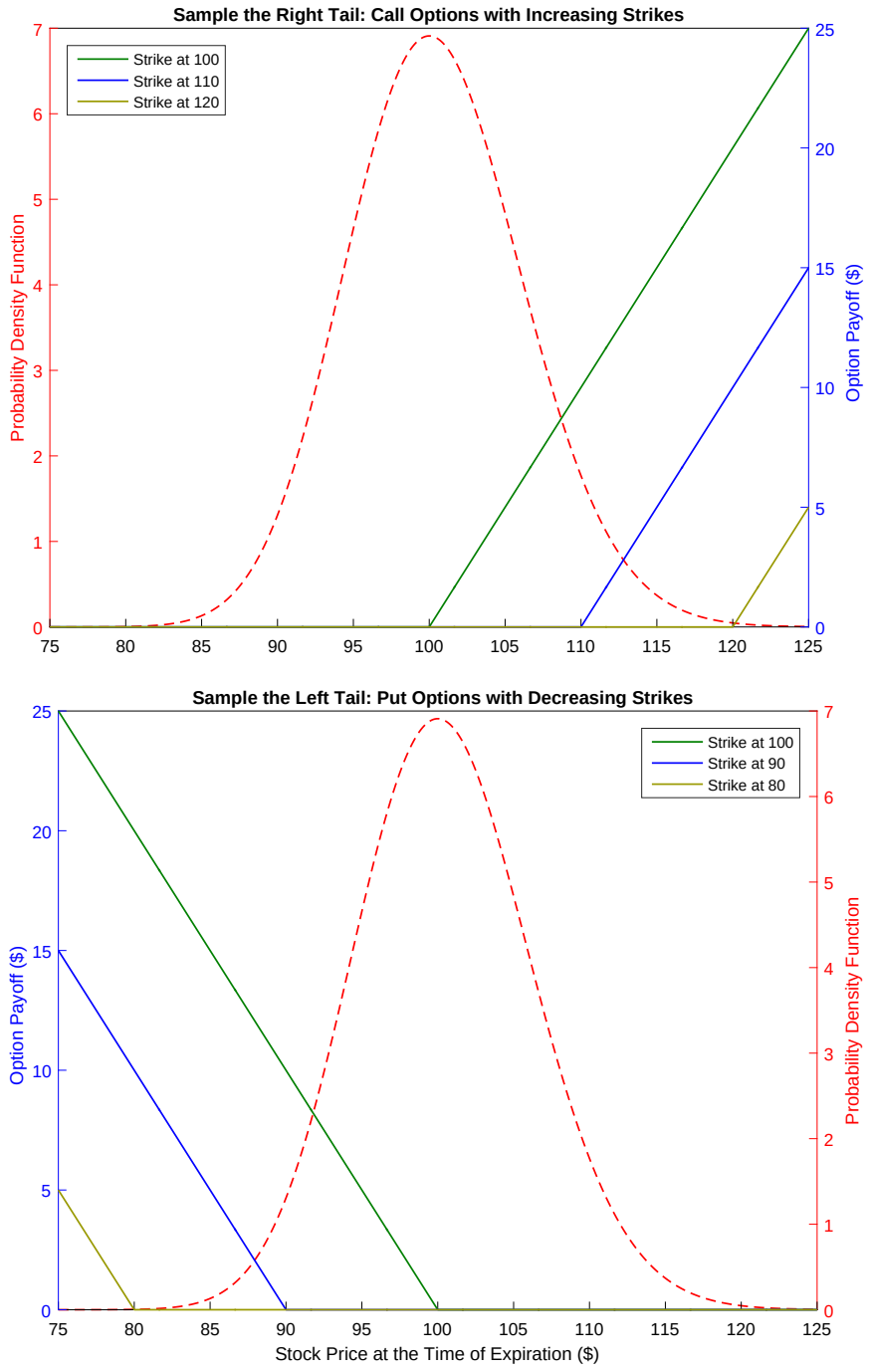


Figure 1: The distribution of a stock plotted against the payoff function of call and put options with varying strike prices.

Let me expand on this example further. As illustrated in Figure 1, moving the strike price of a call option from left to right with increasing strike prices, we are making the call option more and more out of the money. At the same time, this call option becomes more and more sensitive to the right tail of the distribution. Likewise, moving the strike price of a put option from right to left with decreasing strike prices, we are making the put option more and more out of the money. At the same time, this put option becomes more and more sensitive to the left tail of the distribution. Effectively, the market's valuations of such OTM call and put options provide us information about the right and left tails. As we learned early, the left and right tails are not abstract concepts. They are made of extreme financial events: crises show up on the left tail and rallies add to the right tail.

This is as if we are given a high definition camera with a super strong zooming ability. We can point our camera to the right tail and zoom into that area using an OTM call option. Likewise, an OTM put option allows us to zoom into the left tail. If you a photographer, you would be overjoyed to own such a high-definition camera. Likewise, if you are in the business of risk, you would naturally be drawn to these new financial instruments.

- **History:** These new instruments called options first showed up as an exchange-traded product in April 1973, exactly one month before the publication of the Black-Scholes paper. On the first day of trading, 911 contracts of calls were traded on 16 underlying stocks. One option contract is on 100 underlying shares.

  By 1975, the Black-Scholes model was adopted for pricing options. This is an excerpt from an interview with Prof. Merton: *Within months they all adopted our model. All the students we produced at MIT, I couldn't keep them in-house; they were getting hired by Wall Street. Texas Instruments created a specialized calculator with the formula in it for people in the pits. Scholes asked if we could get royalties. They said, "No." Then he asked if we could get a free one, and they said, "No."*

  It was not until 1977, four years after the trading of call options, when trading in put options begins. In 1983, the first index option (OEX) begins trading and a few months later SPX, options on the S&P 500 index, was launched. My PhD thesis was on option pricing and when I first started to work on the CBOE data in 1997, OEX, options on the S&P 100 index, still had a large market presence. By now, it has only a tiny market share. In 1993, CBOE started to publish the VIX index, which was effectively the Black-Scholes implied volatility for an at-the-money one-month to expiration SPX.

In 2004, CBOE launches futures on VIX and later options on VIX.

- **Trading Volume and Market Size:** To gauge the activity of a market, the most frequently used measure is trading volume. For the U.S. equity market, the exchange-listed stocks are traded on 11 stock exchanges ("lit" markets) and about 45 alternative trading systems ("dark pools"). According to summary data from BATS, for the month of September 2015, the average daily trading in the stock market is 7.92 billion shares and $321 billion (dollar volume). For the same month, the overall daily trading volume in the options market is about 16.94 million contracts and $6.30 billion (dollar volume). As you can see, in terms of trading volume, the options market is small compared with its underlying stock market.

  In comparing the trading volumes in the stock and options markets, one interesting observation is that, after the 2008 crisis, the trading in the stock market has been badly hurt. For example, the average NYSE group trading volume peaked around 2.6 billion shares per day in 2008 and has decreased quite dramatically to a level near 1.0 billion shares per day in 2013 and 2014. This is not an NYSE specific problem. The overall stock market trading peaked in 2009 around 9.76 billion shares per day and bottomed to 6.19 billion shares in 2013. By contrast, the trading volume in options did not suffer this dramatic reduction. The average daily trading volume was around 14 million contracts in 2008, increased to 18 million contracts in 2011, and held up steady at around 16 million contracts in 2013.

  In terms of size, the U.S. equity market has a total market value of $26 trillion by end-2014. At the end of September 2015, the open interest for equity and ETF options is 292 million contracts, and 23.7 million contracts for index options. Given that the average premium is around $200 for equity and ETF options and $1,575 for index options, this open interest amounts to $95.7 billion in total market value. Again, the options market is small compared to its underlying stock market.

- **Leverage in Options:** Although the options market is small compared to the underlying stock market, the risk in this market is anything but small. Because of the non-linearity, the leverage inherent in options could be large. Given an investment of the same dollar amount, the profit and loss in options could be many times larger than those in the underlying stocks.

  For example, let's consider a one-month at-the-money put option. Using the Black-Scholes pricing formula, Figure 2 plots the returns to this option as a function of the underlying stock returns, assuming the stock return volatility is 20% per year. As we

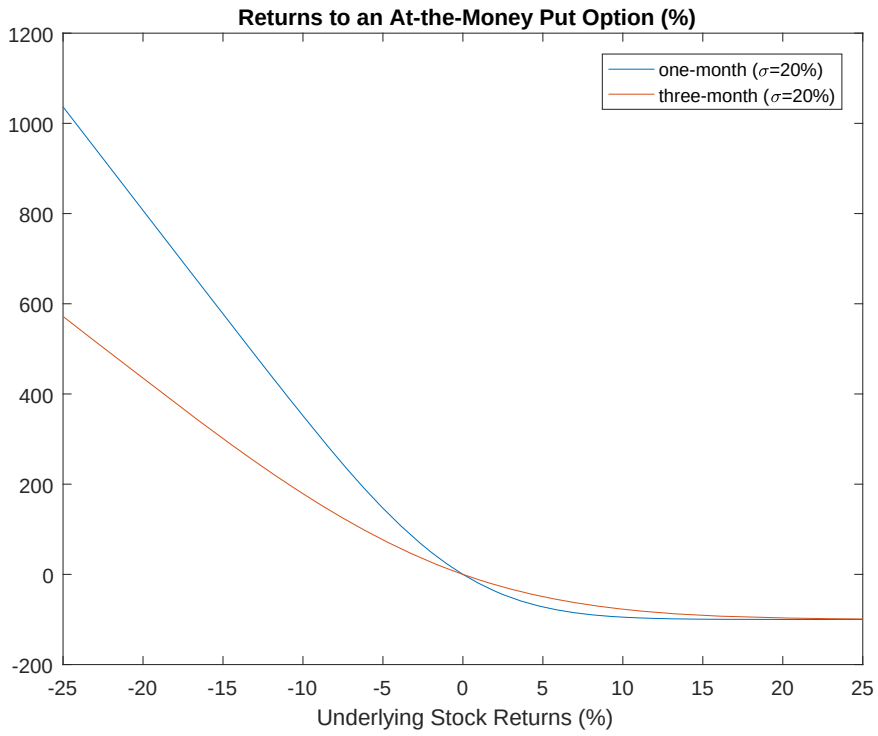


Figure 2: The return of an at-the-money put option plotted against the underlying stock return

can see from the plot, for a 10% drop in the underlying stock price, the option yields a return over 300%. So the inherent leverage in options amplifies a dollar's investment in the underlying stock to 10 dollars in options. Likewise, a 10% increase in the underlying stock price translates to a near -100% drop in the put option. This amplification effect shows up in call options as well, except that the profit and loss of a call option is in the same direction as the underlying stock. Because of these amplification effects, the beta of options on the S&P 500 index can be easily around 20 or -20. Searching through the thousands of stocks listed on the three major U.S. exchanges, you will not be able to find one single stock with this kind of beta. This is what a very simple, almost innocent, non-linearity in the payoff function does to the transformation of risk.

- **Types of Options:** Broadly speaking, there are three types of exchange-traded options: equity, ETF, and index options. Equity options are American-style call and put options on individual stocks. One contract is on 100 underlying shares and the option settles by physical delivery. This CBOE link gives the exact specifications of equity options. Using the September 2015 numbers as an example, the average daily trading volume for equity options is around 7.57 million contracts and $1.58 billion per day in dollar trading volume.

  On any given day, there are thousands of stocks with options traded. Larger stocks usually have higher options trading volume. In September 2015, options on AAPL are by far the most active options traded. Other popular stocks include FB, BAC, NFLX, and BABA, although there is quite a bit of variation over time in terms which stock options show up among the actively traded. If you are curious, this OCC link provides monthly summaries of all equity and ETF option trading volume by exchange.

  ETF options are American-style call and put options on ETFs. Again, one contract is on 100 underlying shares and the option settles by physical delivery of the underlying ETF. This CBOE link gives the exact specifications of ETF options. Since the mid-2000s, the growth in ETF options is an important development in the options market. For September 2015, the average daily trading volume for ETF options is around 7.33 million contracts per day, on par with the trading activity for equity options. The dollar trading volume in ETF options averages to $1.50 billion per day, similar in magnitude to equity options.

  Among the popular ETFs are SPY, EEM, IWM, and QQQ, which command relatively high option trading volume. By far, the most actively traded ETF option is SPY (options on SPDR). For September 2015, SPY options are traded on 12 options exchanges

with an average daily volume of 3.3 million contracts.

Index options are European-style call and put options on stock indices. Except for mini products, one contract is on 100 underlying index. Instead of physical delivery, the settlement of index options is done by cash. This CBOE link gives the specifications of SPX, the most important index options. For September 2015, the average daily trading volume of SPX is about 1.14 million contract and $2.78 billion in dollar trading volume. Recall that the overall dollar trading volume in the options market is about $6.30 billion. This implies that over 30% of the options dollar trading volume comes from SPX. It is therefore not surprising that all options exchanges would like to get involved with this product. So far, CBOE is able to maintain the exclusive license in this product.

You might also notice that both SPX and SPY are trading on the S&P 500 index. There are, however, a few differences between these two products. SPX is on the index itself while SPY is on the ETF SPDR, which is about 1/10 of the index. As a result, per contract, SPX is larger in size than SPY. Recall that average daily trading volume in SPY from 12 exchanges adds up to 3.3 million contracts. This translates to a daily trading volume around $804 million, a large number for ETF and equity options but small compared with the daily dollar volume of $2.78 billion for SPX. Finally, while SPY is an American-style option, SPX is European-style; SPY is physical settlement while SPX is cash settlement.

Regardless of their differences, SPX and SPY share the same underlying. Therefore there must be market participants who actively trade between these two contracts to profit from any temporary mis-pricing between the two. As a result, the pricing of these two contracts should be very much aligned with one another, taking into account of the difference in their exercise style. For those who are interested, it might be a good exercise to go to the CBOE's website to get quotes for near-the-money near-the-term SPX and SPY call and put options, back out the Black-Scholes implied volatilities from these contracts and see if there are any significant pricing differences (above and beyond the quoted bid and ask spreads).

- **Options Exchanges:** As we see earlier, over 30% of the option dollar trading volume comes from SPX: call and put options on the S&P 500 index. Not surprisingly, CBOE fought really heard to keep its exclusive rights to SPX. In 2012, after 6 years of litigation, CBOE won the battle and was able to retain its exclusive licenses on options on the S&P 500 index. As a result, CBOE remains its dominance in index options with

over 98% of the market share. In addition to SPX, options on VIX have also grown in popularity, which is also traded exclusively on CBOE.

In other areas, however, CBOE has not been able to retain its market power. Until the late 1990s, CBOE was the main exchange for options trading. By the early 2000s, however, CBOE was losing its market share in equity options to new option exchanges like ISE. For equity options in September 2015, CBOE accounts for 16.32% of the trading volume, PHLX has a market share of 17.50%, and the rest are shared by BATS (14.37%), ARCA (11.81%), ISE (10.53%), AMEX (8.89%) and others. Trading in ETF options took off around the mid-2000 and have been spread over many options exchanges in a way similar to equity options: CBOE (15.93%), ISE (15.55%), PHLX (15.02%), BATS (10.48%), ARC (10.15%), AMEX (10.14%), and others.

You might have noticed the fragmentation of the options market. Indeed, equity and ETF options are traded in 12 different options exchanges. This phenomenon of market fragmentation is not option specific. For example, US stocks regularly trade on 11 exchanges. In addition to these exchanges which are called "lit" markets, a non-trivial amount (20% to 30% in 2015) of stock trading is done in alternative trading systems such as "dark pool."

- **Market Participants:** One advantage of options being traded on exchanges is its accessibility. Investors of all types come to the market to trade. Another advantage is its transparency. Information on transaction prices and volumes is readily available to investors. On any given day, you can see how many put options are bought on the S&P 500 index or on AAPL versus how many call options are bought. The same thing cannot be said about the over-the-counter (OTC) derivatives market. While pricing information on OTC derivatives can be obtained from Bloomberg or Datastream, the real-time transaction information is very much protected by dealers as proprietary information. In my personal view, if the trading information in products such as CDS, CDO, synthetic CDO, and CDO2 were available to the public back in 2005, more people would have paid attention to this market.

  Like most markets, there are designated market makers in the options market. Their presence in the market is to facilitate trading and provide liquidity. They make money by quoting bid and ask prices: buy at the bid and sell at the ask. The bid-ask spread (ask price - bid price) is the source of their profit. In the options market, the percentage bid-ask spread is much larger than that of the underlying stock, reflecting the leverage risk inherent in options. It is also a reflection of the relative illiquidity in options.

When there are buying and selling imbalances, market makers might have to keep an inventory, which exposes them to market risk. This risk exposure is further exaggerated if this imbalance is caused by some private information the market maker is not aware of (information asymmetry). The inventory cost and the cost of information asymmetry are two important drivers for the bid and ask spread in financial prices. In the options market, it is typical for a market maker to minimize his exposure to the underlying stock by delta hedging.

Coming back to the topic of SPX and SPY, two options products with very similar underlying risk. You might notice that there is a substantial difference in their bid/ask spreads. In particular, the average bid/ask spreads (as a percentage of the option price) are much higher for SPX than SPY. The average percentage bid/ask spread for SPX is about 9% while that for SPY is about 1%. If the market risk is similar, then where does this difference in trading cost arise?

Investors who trade against the market makers can be summarized into four groups: customers from full service brokerage firms (e.g., hedge funds), customers from discount brokerage firms (e.g., retail investors), and firm proprietary traders. Using CBOE data from 1990 through 2001, we see that customer from full serve brokerage firms are the most active participants in the options market while firm proprietary traders concentrate their trading mostly on index options as a hedging vehicle. Of course, these are older data and the options market has exploded after 2001.

Another way to look at the market participants is through their trading activities against the market makers. Some investors come to the options market to buy options to open a new position, while other buy options to close an existing position. Some sell options to new a new position while others sell options to close an existing position. In doing so, their trading motives are very different.

# 2 The Black-Scholes Option Pricing Model

- **The Model:** Let $S_t$ be the stock price at time $t$. For simplicity, let's first assume that this stock pays no dividend. Later we will add dividend back. We model the dynamics of the stock price by the following model (geometric Brownian motion):

$$dS_t = \mu\, S_t\, dt + \sigma\, S_t\, dB_t\,. \tag{1}$$

This equation does not look very appealing at the moment, but you will come to

appreciate or even like it later. Under this model, the expected stock return is $\mu$ and its volatility is $\sigma$, both numbers are in annualized terms. So if you like, $\mu$ is about 12% and $\sigma$ is about 20%. Moreover, under this model, stock returns (to be more precise, log-returns) are normally distributed. Let me use the rest of this section to explain why it is so.

Let $S_T$ be the stock price at time $T$. Implicitly we are planning ahead for the time $T$, when the option expires. Standing here at time 0 and holding a European-style option, all we care about is the final payoff:

$$\text{Payoff of a call option struck at K} = (S_T - K)\,\mathbf{1}_{S_T>K} \tag{2}$$

where $\mathbf{1}_{S_T>K} = 1$ if $S_T > K$ and zero otherwise. Let's focus on call options for now. Once we now how to deal with call options, the put/call parity will get us to put options very easily.

Option pricing bolts down to calculating the present value of the payoff in equation (2). How should this calculation be done? What is the discount rate to use in order to bring the random cash flow to today? Let's keep this question hanging for a while.

- **Brownian Motion:** Since it is the first time we are working with Brownian motions, let me summarize the following three important properties of Brownian motions and relate them to Finance:
  - *Independence of increments:* For all $0 = t_0 < t_1 < \ldots < t_m$, the increments are independent: $B(t_1) - B(t_0)$, $B(t_2) - B(t_1)$, $\ldots$, $B(t_m) - B(t_{m-1})$ . Translating to Finance: stock returns are independently distributed. No predictability and zero auto-correlation $\rho = 0$.
  - *Stationary normal increments:* $B_t - B_s$ is normally distributed with zero mean and variance $t-s$. Translating to Finance: stock returns are normally distributed. Over a fixed horizon of $T$, return volatility is scaled by $\sqrt{T}$.
  - *Continuity of paths:* $B(t)$, $t \geq 0$ are continuous functions of $t$. Translating to Finance: stock prices move in a continuous fashion. There are no jumps or discontinuities.
- **The Model in $R_T$:** Let's perform this very important transformation:

$$S_T = S_0\, e_T^R\,.$$

Another way to look at it is by,

$$R_T = \ln(S_T) - \ln(S_0)\,,$$

which tells us that $R_T$ is the log-return of the stock over the horizon $T$. Now I am going to do one magic and you just have to trust me on this. Next semester when you take 450, you will learn the mechanics behind it, which is call the Ito's Lemma.

$$dR_t = \left(\mu - \frac{1}{2}\sigma^2\right)\,dt + \sigma\,dB_t$$

Comparing with equation (1), the dynamics of $R_t$ is simpler. It does not have those $\mu S_t$ and $\sigma S_t$ terms. Instead, we have $\mu - \sigma^2/2$ as its drift and $\sigma$ as its diffusion coefficient. The extra term of $\sigma^2/2$ is often call the Ito's term.

With this dynamics for $R_t$, we can now fix the time horizon $T$ and write out $R_T$:

$$R_T = \int_0^T dR_t = \left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}\,\epsilon_T\,, \tag{3}$$

where $\epsilon_T$ is a standard normal random variable (zero mean, variance equals to 1). You will agree with me that $\int_0^T dt = T$. Let me explain why $\int_0^T dB_t = B_T - B_0$ is $\sqrt{T}\,\epsilon_T$: it comes from the second property, stationary normal increments, of the Brownian motion.

When it comes to valuation under the Black-Scholes model, the math will be done at the level of equation (3). As you can see, it is not that scary, isn't it? This model tells us that the log-return of a stock over a fixed horizon of $T$ is normally distributed with mean $(\mu - \sigma^2/2)\,T$ and standard deviation of $\sigma\sqrt{T}$. Other than the Ito's term, $\sigma^2/2$, everything looks quite familiar. No?

- **The Ito's Term:** Now let me explain why we have this Ito's term. In the continuous-time model of equation (1), the stock price grows at the instantaneous rate of $\mu\,dt$:

$$E(S_T) = S_0\,e^{\mu T}\,,$$

or equivalently, with a continuously compounded discount rate $\mu$:

$$S_0 = e^{-\mu T}\,E(S_T)\,.$$

Now let's do the same calculation with our model for log-return in Equation (3),

$$E(S_T) = S_0\, E\left(e^{R_T}\right) .$$

When it comes to calculating expectation of a convex function involving a normally distributed random variable $x$, this is a useful formula for you to have

$$E(e^x) = e^{E(x)+\mathrm{var}(x)/2} .$$

Let me emphasize, this works only when $x$ is normally distributed. Applying this formula to the above calculation, we have

$$E(S_T) = S_0\, E\left(e^{R_T}\right) = S_0\, e^{E(R_T)+\mathrm{var}(R_T)/2} = S_0\, e^{\left(\mu-\sigma^2/2\right)T+\sigma^2\, T/2} = S_0\, e^{\mu T} ,$$

which is exactly what we wanted in the first place.

To summarize, the transformation from $S_T$ to $\ln(S_T) - \ln(S_0)$ introduces some concavity, because $\ln(x)$ is a concave function. This is why $-\sigma^2/2$ shows up in $R_T$. The transformation from $R_T$ to $e^{R_T}$ introduces some convexity, because $e^x$ is a convex function, and $\sigma^2/2$ gets added back during the transformation. So everything works out.

In essence, Mr. Ito is busy because we are doing concave/convex transformations on random variables. If there is no random variable involved, then Mr. Ito will not be this busy. For example, let's make $x$ a number by setting $\mathrm{var}(x) = 0$. What do we have for $E(e^x) = e^{E(x)+\mathrm{var}(x)/2}$? We have $E(e^x) = e^x$ and nothing else. The Ito's term disappeared.

- **Risk-Neutral Pricing:** Now let's come back to the present value calculation. As discussed earlier, the payoff of a call option at time $T$ is as in Equation (2). It is a random payoff, depending on the realization of $S_T$. It is a non-linear random payoff with a kink at the strike price $K$: the payoff is zero if $S_T$ falls below $K$ and is $S_T - K$ if $S_T$ rallies above $K$ and the option is exercised at time $T$. So what is the present value of this random non-linear payoff? Which discount rate should we use?

  Risk-neutral pricing is the answer to that question. Although it has "risk-neutral" in its name, it is anything but risk-neutral. Let me first tell you the approach of the

risk-neutral pricing. Recall that after some hard work, we have

$$R_T = \left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma\sqrt{T}\,\epsilon_T\,.$$

I am going to call this model the actual dynamics and label it by "P." Then I am going to introduce a different model, called risk-neutral dynamics and label it by "Q."

$$\text{Actual Dynamics ("P"):} \qquad R_T = \left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma\sqrt{T}\,\epsilon_T \tag{4}$$

$$\text{Risk-Neutral Dynamics ("Q"):} \quad R_T = \left(r - \frac{1}{2}\sigma^2\right) T + \sigma\sqrt{T}\,\epsilon_T^{\mathrm{Q}} \tag{5}$$

By writing down the model in the Q-dynamics, I am bending the reality by forcing the stock return to grow at the riskfree rate $r$. And then I am going to do my present value calculation under this bent reality: 1) the expectation of the future cash flow is done under the Q-measure and 2) this expectation is discounted back to today using the riskfree rate $r$. And somehow, two wrongs make one right, the calculation works out. You just have to trust me on this. This pricing framework is widely adopted on Wall Street in fixed income, credit, and options.

- **Pricing a Stock:** Before applying this risk-neutral pricing framework on options , let's first try it on something easier: the linear random payoff of $S_T$. We know what the answer should be: the present value should be $S_0$. We've already done it under the P-dynamics: $S_0 = e^{-\mu T} E(S_T)$. It works out and using $\mu$ as the discount rate makes perfect sense ... because this is how the dynamics is written.

  Now let's do it under the Q-dynamics:

$$e^{-rT}\,E^{Q}(S_T) = e^{-rT}\,S_0\,e^{rT} = S_0\,.$$

  So it also works! Just to emphasize that risk-neutral pricing has nothing to do with investors being risk-neutral, let's bring in a risk-neutral investor to price the same stock. He takes the P-dynamics (because it is the reality) and discounts the cash flow with riskfree rate $r$ (because he is risk neutral):

$$e^{-rT}\,E^{\mathrm{P}}(S_T) = e^{-rT}\,S_0\,e^{\mu T} = S_0\,e^{(\mu - r)T}$$

  So he is paying more than $S_0$ for the same cash flow. Why? Because he is risk-neutral. Recall that if $S_T$ is the market portfolio, then $\mu - r$ is the market risk premium.

Risk-averse investors demand a premium for holding the systematic risk in the market portfolio. That gives rise to the positive risk premium in $\mu - r$. A risk-neutral investor, however, is not sensitive to risk. As such, he is willing to pay more for the stock.

This exercise might seem trivial mathematically, but it is very useful in clearing our thoughts. In particular, I would like to emphasize that risk-neutral pricing does not mean pricing using a risk-neutral investor. In a way, this name "risk-neutral pricing" is unfortunate and confusing.

- **Pricing the Option:** We are now ready to price the option. Let $C_0$ be the present value of a European-style call option on $S_T$ with strike price $K$:

$$C_0 = e^{-rT} E\left(\left(S_T - K\right) \mathbf{1}_{S_T > K}\right) = e^{-rT} E^Q\left(S_T \mathbf{1}_{S_T > K}\right) - e^{-rT} K E^Q\left(\mathbf{1}_{S_T > K}\right)$$

Now let me cheat a little by going directly to the solution,

$$C_0 = S_0\, N(d_1) - e^{-rT} K N(d_2)\,,$$

where $N(d)$ is the cumulative distribution function of a standard normal x:

$$N(d) = \text{Prob}\left(x \leq d\right) = \int_{-\infty}^{d} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx\,.$$

In Matlab, $N(d)$ is normcdf(d). The two critical values $d_1$ and $d_2$ are,

$$d_1 = \frac{\ln\left(S_0/K\right) + \left(r + \sigma^2/2\right) T}{\sigma\sqrt{T}}\,; \quad d_2 = \frac{\ln\left(S_0/K\right) + \left(r - \sigma^2/2\right) T}{\sigma\sqrt{T}}$$

Comparing where we are now with the solution, we see some internal logic. In particular, it is obvious that

$$N\left(d_2\right) = E^Q(\mathbf{1}_{S_T > K}) = \text{Prob}^Q\left(S_T > K\right)$$

and

$$N\left(d_1\right) = e^{-rT} E^Q\left(\frac{S_T}{S_0} \mathbf{1}_{S_T > K}\right)\,.$$

- **Understanding $N(d_2)$ in the Black-Scholes formula:** The part associated with $N(d_2)$ is actually pretty easy. It calculates the probability that the call option is in

the money under the Q-measure. So let's work it out:

$$\text{Prob}^Q\left(S_T > K\right) = \text{Prob}^Q\left(S_0\, e^{R_T} > K\right) = \text{Prob}^Q\left(e^{R_T} > K/S_0\right) = \text{Prob}^Q\left(R_T > \ln(K/S_0)\right) ,$$

where, in the last step, I took a log on both side of the inequality, which is OK because $\ln(x)$ is a monotonically increasing function in $x$.

Now let's use the Q-dynamics of $R_T$ in Equation (5) to get,

$$\text{Prob}^Q\left(R_T > \ln(K/S_0)\right) = \text{Prob}^Q\left(\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}\,\epsilon_T^{\text{Q}} > \ln(K/S_0)\right) .$$

Moving things left and right, we get

$$\text{Prob}^Q\left(\epsilon_T^Q > \frac{-\ln(S_0/K) - \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right) ,$$

or equivalently,

$$\text{Prob}^Q\left(-\epsilon_T^Q < \frac{\ln(S_0/K) + \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right) ,$$

which is really $N(d_2)$, knowing that $\epsilon_T^Q$ is standard normally distributed.

- **Understanding $N(d_1)$ in the Black-Scholes formula:** The part associated with $N(d_1)$ is more subtle. Recall that

$$N\left(d_1\right) = e^{-rT} E^Q\left(\frac{S_T}{S_0}\,\mathbf{1}_{S_T>K}\right)$$

So $N(d_1)$ involves a calculation that takes into account that we are calculating the expectation of $S_T$ when $S_T$ is greater than K (the option expires in the money). So it is not a simple probability calculation such as $N(d_2)$. Here, it involves an interaction term. As a result $N(d_1)$ should always be larger than $N(d_2)$. This is true because $d_1 = d_2 + \sigma\sqrt{T}$. As you will see later, this difference between $d_1$ and $d_2$ is really where the option value of an option comes from. In other words, $\sigma\sqrt{T}$ is the best summary of the option value.

Given the amount of math we have been doing up to this point, I have a feeling that most of you are not willing to go further. For those of you who are interested, you can do the math to prove that $N(d_1)$ is in fact $e^{-rT}E^Q\left(\frac{S_T}{S_0}\,\mathbf{1}_{S_T>K}\right)$.

For those who are not willing to go through the math, let me offer this observation.

Under the Q-dynamics, the drift in $R_T$ is $(r-\sigma^2/2)T$ and the volatility is $\sigma\sqrt{T}$. That's how we get the expression of $d_2$ (and our previous calculation just proved this point). Comparing $d_1$ and $d_2$ this way, we notice that suppose we bend the reality further by making the drift in $R_T$ to be $(r+\sigma^2/2)T$ and keep the same volatility. Then, under this strange dynamics, let's call it $QQ$, we have $N(d_1) = \text{Prob}^{QQ}(S_T > K)$. Intuitively, because of the interaction term, the valuation is higher. One simple way to express this higher valuation is by allowing $R_T$ to grow faster than its Q-measure, with a drift of $(r+\sigma^2/2)T$. Under this probability measure, the probability of $S_T$ is greater than $K$ (the option expires in the money) becomes $N(d_1)$. I'll stop here.

- **Add Dividend Yield:** We are going to apply the Black-Scholes model to SPX. So it is important that we can handle stocks paying dividend with a constant dividend yield, which, for the S&P 500 index, is a good enough approximate. Let $q$ be the dividend yield. Again, let $S_T$ be the time-$T$ stock price, ex dividend. Then, the stock dynamics becomes,

$$dS_t = (\mu - q)\ S_t\, dt + \sigma\, S_t\, dB_t\,.$$

And the dynamics for $R_T$ changes to

$$\text{Actual Dynamics ("P"):} \qquad R_T = \left(\mu - q - \frac{1}{2}\sigma^2\right) T + \sigma\sqrt{T}\,\epsilon_T$$

$$\text{Risk-Neutral Dynamics ("Q"):} \quad R_T = \left(r - q - \frac{1}{2}\sigma^2\right) T + \sigma\sqrt{T}\,\epsilon_T^{\text{Q}}$$

And the Black-Scholes pricing formula becomes

$$C_0 = e^{-qT}\, S_0\, N(d_1) - e^{-rT}\, K\, N(d_2)\,,$$

where $N(d)$ is the cumulative distribution function of a standard normal and

$$d_1 = \frac{\ln(S_0/K) + (r - q + \sigma^2/2)\, T}{\sigma\sqrt{T}}\,; \quad d_2 = \frac{\ln(S_0/K) + (r - q - \sigma^2/2)\, T}{\sigma\sqrt{T}}$$

- **Arbitrage Pricing and Dynamic Replication:** In Finance, when it comes to valuation, there are just two approaches: equilibrium pricing and arbitrage pricing. We've touched on equilibrium pricing in the CAPM, where mean-variance investors optimize their utility functions and the equity and bond markets clear. What we've been doing so far in this class falls squarely into the category of arbitrage pricing. The essence of arbitrage pricing is replication: replicate a stream of random payoffs with

existing securities whose market values are known to us. The present value of this cash flow equals to the cost of the replication.

The best example in our current setting is the put/call parity. As I am sure that you've learned in 15.415 (or 15.401), the time-$T$ payoff of buying a European-style call and selling a European-style put (with the same strike price $K$) is the same as taking a long position in the underlying stock and borrowing $K$ from the bond market. The present value of the underlying stock is $e^{-qT}S_0$, where, as usual, we use ex dividend stock price. The present value of the bond-borrowing portion is $e^{-rT}\,K$, with $r$ being the riskfree rate, continuously compounded. So the replication cost is $e^{-qT}S_0 - e^{-rT}\,K$. The present value of buying a call and selling a put is, by definition, $C_0 - P_0$. As a result, $C_0 - P_0 = e^{-qT}S_0 - e^{-rT}\,K$.

As you can see, in getting this relation, we do not have to use any model, just simple logic. In practice, this put/call ration holds pretty well in the market. There are investors actively arbitrage between the options and the cash (i.e., the S&P 500 index or the S&P 500 index futures via "E-mini") markets. Even if the Black-Scholes model fails (which it does), this relation still holds. Arbitraging using put/call parity is very similar to arbitraging between the futures and cash markets (arbitraging between Chicago and New York).

When it comes to pricing call and put options, however, we do need to use a model. So far, we've used the Black-Scholes model. It turns out that even with a stock and a bond, we can still replicate the non-linear payoff of an option. This is the important insight of Prof. Black, Merton, and Scholes: dynamic replication. You need to continuously rebalance your hedging portfolio, doing delta hedging at a super high frequency. I am sure that you've got a heavy dosage of that in your 15.415. So I am not going to spend time on dynamic replication or delta hedging.

Recall that the third property of a Brownian motion is continuity of paths. This implies that stock prices move in a continuous fashion. There is no jumps or discontinuities. This is why models like geometric Brownian motions are called diffusion models. As you can see, the property of dynamic replication falls apart as soon as we move away from the Brownian motion by adding random jumps to the model. This is just one example. If we add another streams of random shocks to volatility, making it a stochastic process (instead of a number $\sigma = 20\%$), then this replication also falls apart.

As such, the Black-Scholes formula is very much confined to the model itself. We will see that the Black-Scholes model does not hold very well in the market. We will then

extend in two dimensions: adding jumps to the model to allow crashes; relaxing $\sigma$ from a number to a stochastic process and build a stochastic volatility model.

- **Why so many equations?** Since Fall 2015, because of the MFin students, I made a conscious effort in being as rigorous as possible and giving you as much detail as possible. While using the Black-Scholes model as a black box is fine for most people, I feel that most of you deserve to know a little bit better. In past years, 15.450 was taught along with 15.433. So I made the comfortable choice of letting the professor in 15.450 carry more of the math burden. Now that 15.450 has been moved to the Spring semester, I feel that I've lost my excuse. And Prof. Wang kept asking me to push you more. So this is my effort in pushing you.

  If you've seen this before, don't presume that you know everything. Honestly, I started to work in this area as soon as I entered the PhD program at Stanford GSB 20 years ago. But I've only developed these intuitions over the years. So take your time to digest the materials and make them your own.

# 3 Using the Black-Scholes Formula

- **Pricing ATM Options:** By definition, an at-the-money option has the strike price of $K = S_0\, e^{(r-q)T}$. Going back to $d_1$ and $d_2$, we notice that by setting the strike price at this level, $d_1 = \frac{1}{2}\sigma\sqrt{T}$ and $d_2 = -\frac{1}{2}\sigma\sqrt{T}$. Effectively, by having an option with this strike price, we take away the moneyness component of the option and focus exclusively on the option value. Also notice that at this strike price, $e^{-qT}\, S_0 = e^{-rT}\, K$, which implies that, via the put/call parity, $C_0 = P_0$ for this pair of at-the-money call and put options. For the case of $\sigma = 20\%$ and $T = 1/12$, we have $d_1 = \sigma\sqrt{T}/2 = 0.0289$. Figure 3 plots the respective $N(d_1)$ and $N(d_2)$ for the case.

  Applying the Black-Scholes formula, we have

$$C_0 = P_0 = S_0\left(N(d_1) - N(d_2)\right) = S_0\left[N\left(\frac{1}{2}\sigma\sqrt{T}\right) - N\left(-\frac{1}{2}\sigma\sqrt{T}\right)\right].$$

  Using the fact that $N(d)$ is the cdf of a standard normal:

$$N(d) = \int_{-\infty}^{d} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx\,,$$

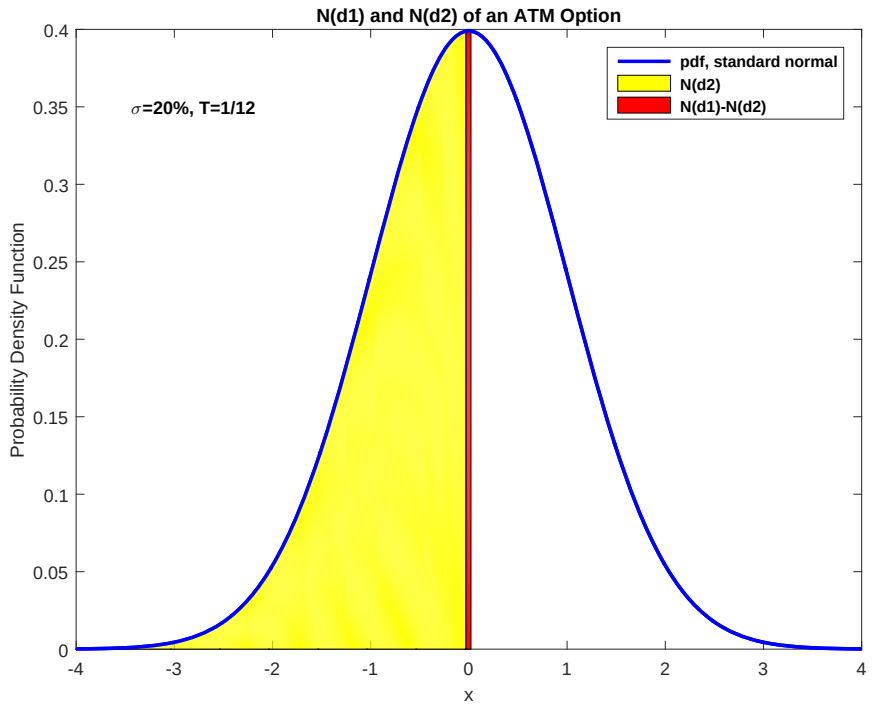


Figure 3: The $N(d_1)$ and $N(d_2)$ for an one at-the-money call or put option with one-month to expiration. The underlying stock volatility is 20%.

we can further simplify the pricing formula,

$$\frac{C_0}{S_0} = \frac{P_0}{S_0} = \frac{1}{\sqrt{2\pi}} \int_{-\frac{1}{2}\sigma\sqrt{T}}^{\frac{1}{2}\sigma\sqrt{T}} e^{-\frac{x^2}{2}}\, dx\,.$$

Now let's use a Taylor expansion that is very useful in Finance: $e^x \approx 1 + x$, for small $x$. Applying this to the integrand,

$$e^{-\frac{x^2}{2}} = 1 - \frac{x^2}{2}\,.$$

Let's replace the integrand with this approximate:

$$\frac{C_0}{S_0} = \frac{P_0}{S_0} \approx \frac{1}{\sqrt{2\pi}} \int_{-\frac{1}{2}\sigma\sqrt{T}}^{\frac{1}{2}\sigma\sqrt{T}} \left(1 - \frac{x^2}{2}\right) dx = \frac{1}{\sqrt{2\pi}} \left(\sigma\sqrt{T} - \frac{1}{24}\left(\sigma\sqrt{T}\right)^3\right) \approx \frac{1}{\sqrt{2\pi}}\sigma\sqrt{T}\,,$$

where I dropped the cubic term to make our approximation even simpler. But you can see, if you include the next order of approximation, the net effect will make the option price lower. This approximation works well for small $\sigma\sqrt{T}$. For a typical one-month

option on the S&P 500 index, $\sigma = 0.20$ and T=1/12, we have $\sigma\sqrt{T}$ being around 0.0577. As a comparison, the higher order term $\left(\sigma\sqrt{T}\right)^3/24$ is $8 \times 10^{-6}$. So this level of $\sigma\sqrt{T}$, our approximation works really well.

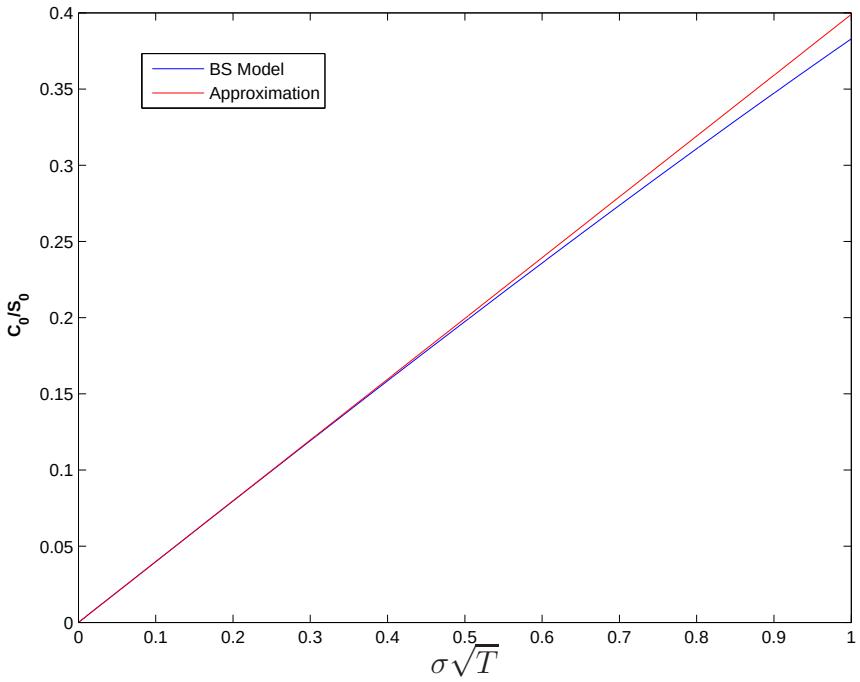


Figure 4: The ratio of an at-the-money call or put option price to the underlying stock price, $C_0/S_0$ or $P_0/S_0$, as a function of $\sigma\sqrt{T}$. The approximation of $C_0/S_0 = P_0/S_0 \approx \sigma\sqrt{T}/\sqrt{2\pi}$ is in red and the Black-Scholes pricing is in blue.

As shown in Figure 4, as $\sigma\sqrt{T}$ becomes large, this approximation becomes imprecise. Moreover, the approximation is bias upward compared with the Black-Scholes pricing. This makes sense because the next higher order term is negative. It also makes sense because $C_0/S_0$ cannot grow linearly with $\sigma\sqrt{T}$ forever. The call option price is bounded from above by the underlying stock price: $C_0/S_0$ cannot be bigger than 1. At some point, this ratio has to taper off.

What kind of options will give us $\sigma\sqrt{T}$ that is too large for this approximation to work? Options on volatile stocks with long time to expire. For example, for an option with $\sigma = 100\%$ and 1 year to expiration, $\sigma\sqrt{T} = 1$. As you can see from Figure 4, our approximation is no longer very good.

- **ATM Options as Financial Vehicles on $\sigma\sqrt{T}$:** In spending time to analyze the at-the-money options, we learned an important lesson. In fact, it is the cleanest way to

understand what options are really about. By buying a call option, we get a positive exposure to the underlying stock; by buying a put option, we get a negative exposure. Neither of these exposures is unique to options. There are other ways we can get this kind of exposure. And the exposure can be easily hedged out by stocks. But what's unique about options is the volatility exposure. In the Black-Scholes model, volatility is a constant. So you might not appreciate the significance of this volatility exposure. As soon as we allow volatility to move around, which is true in reality, then you find in options a vehicle that is unique in offering exposures to $\sigma\sqrt{T}$. Nothing in the stock market can offer this kind of exposure.

Recall that dynamic replication makes options a redundant security within the Black-Scholes model. At that point, you might be wondering to yourself that: if it is redundant, then what is the point? Well, in reality, with random shocks to volatility and fat-tails in stock returns, options are not at all redundant. That is why, as beautiful and revolutionary as the dynamic replication theory is, I do not want us to spend too much time on it.

Going back to our discussions regarding $N(d_1)$ and $N(d_2)$, the example of ATM options further clarifies what really matters in $d_1$ and $d_2$. It's the fact that $d_1$ is always larger than $d_2$, by the amount of $\sigma\sqrt{T}$ in the Black-Scholes model. If you trace back to the calculation of $d_1$, you notice that it comes from $E^Q\left(S_T\mathbf{1}_{S_T>K}\right)$, the positive interaction between $S_T$ and $\mathbf{1}_{S_T>K}$. Within the Black-Scholes setting, we have the exact formulation of this option value. As we later move away from the Black-Scholes model, $N(d_1)$ and $N(d_2)$ will be replaced by other formulas. That is why I have been emphasizing calculations like $E^Q\left(\mathbf{1}_{S_T>K}\right)$ and $E^Q\left(S_T\mathbf{1}_{S_T>K}\right)$ for call options. These calculations are the main building blocks of a call option, whose values might be different in different models. Likewise, for put options, calculations like $E^Q\left(\mathbf{1}_{S_T<K}\right)$ and $E^Q\left(S_T\mathbf{1}_{S_T<K}\right)$ are the main building blocks.

- **The Black-Scholes Option Implied Volatility:** Once we understand that options are unique financial vehicles for volatility, then volatility will be the first thing we would like to learn from options. Indeed, the Black-Scholes option implied volatility is such a concept.

  For a call option with strike price $K$ and time to expiration $T$, we can calculate its Black-Scholes price by plugging the model parameters. We obtain the underlying stock price $S_0$ from the stock market, the riskfree rate $r$ from the Treasury or LIBOR market. If this option is on the S&P 500 index, we can assume a flow of dividend payment in

the form of a dividend yield $q$. We can approximate $q$ with its historical average, say 2%. Now the only parameter left for us to move around is $\sigma$. Of course, we can go to the underlying stock market to measure the volatility. But let's not do that. Let's instead back out the volatility $\sigma^I$ so that the model price for this option agrees with the market price of this option. This is the Black-Scholes implied volatility.

In doing this exercise, we are not assuming the Black-Scholes model is correct. We are only using the model as a tool for us to transform the option price from the dollar space to the volatility space. Why is this useful? Because options with different strike prices and times to expiration will differ quite a lot in their market value. A deep in-the-money option might be worth hundreds of dollars, while a deep out-of-the-money option on the same underlying might be worth just a few dollars. A short-dated options is worth much less than a long-dated options. Since all of these options are on the same underlying, you would like to be able to compare their pricing. But comparing these options in the dollar space is not at all intuitive. By contrast, all of these options share the same underlying. Hence the same $\sigma$. So comparing these options in the volatility space is much more intuitive and productive. In fact, in OTC markets, options are typically quoted not in dollar but in the Black-Scholes implied volatility. This is analogous to the adoption of yields in the bond market. So Black-Scholes implied vols in options and yields in bonds.

# Appendix

## A Valuation Models in Finance

As we move on to options and fixed-income products, concepts such as present value calculation will take center stage. Looking back, you might have noticed that in our equity classes, we worked almost exclusively in the return space. We analyze the distribution of stock returns, estimate the expected return, investigate the return predictability, and study the various models of return volatility. Very rarely did we talk about valuation. For example, AAPL has a market capitalization of $642B with $112 a share right now. What kind of Finance models do we use to price this stock? Can the same model be used to price other stocks? How well does such a model work in practice?

The one exception was when we work with the book-to-market ratio in the Fama-French model. We use the book value of equity as a benchmark for the market value of equity. If investors think of buying the stock as buying the book value of the firm, then this ratio should be around one. In practice, we noticed a wide range of book-to-market ratios. For example, as of July 2015, the average book-to-market ratio is around 0.095 for stocks in decile 1 and 1.339 for those in decile 10. AAPL with its book-to-market ratio of 0.2 belongs only to decile 2. Conceptually, we can say that stocks with low book-to-market ratios are those with great growth potential. As such, investors are willing to pay multiple (in the case of 0.095, 10 times) of the book value. Quantitatively, however, why some stocks are priced at 10 times while other stocks are priced at 0.75 times? Do we have one good model to give accurate prices to this cross-section of stocks with varying book-to-market ratios?

By now, you've probably been taught various valuation models that combine cash flows with discount rates. You project the future cash flows of a firm or a project and discount them back using some discount rates estimated using a Finance model, say the CAPM. Without a question, these frameworks are useful in helping us think through the key components in a valuation project. But, quantitatively, these models do not offer the kind of precision and rigor as other models in Finance. And in practice, this seems to be true as well.

When I first read the fascinating book on the RJR Nabisco deal, "Barbarians at the Gate," my mouth was wide open as I flipped through the pages. For such a large deal, the valuation seems to be rather flexible. Over the short time span of one month and 11 days, the valuation moved from the initial $17 billion with $75 a share to $24.88 billion with $109 a share. This might be an extreme case, but other books on private equity, for example, "King of Capital," left me with the same impression: there is a lot of flexibility in valuation

in this space. If you look at the venture capital space, where even the projection of future cash flow is very much up in the air, you see a similar pattern. I am not an expert in either of these areas, but it is safe to say that the level of precision required of a Finance model is relatively low in these areas, or the margin of errors allowed for such valuation models is rather high.

In writing this introduction on valuation, my objective was to compare and contrast the role of valuation in various parts of Finance. On the one end of the spectrum, you have valuations in VC and private equity. In this space, the cash flows are highly uncertain; which discount rates to use is also not clear. The role of a valuation model in such a setting is indeed very limited. If you are working in this area, spending time to perfect your Finance model is not at all your number one priority. On the other end of the spectrum, you have valuations in options and fixed income. In options, the cash flow comes from the fluctuation of the underlying stock price. In fixed income, the cash flow comes from the coupon and principal payments. In both cases, the cash flow can be modeled rather precisely and the present value calculation can be done with super high precision. In these areas, people take their valuation models rather seriously. If anything, the danger is that people take their models too literally to the extent that they are lost in their models.

I hope that you do not read this introduction as "one against another." This concern made me move this introduction to the appendix so as not to distract you from the main topic. The role of a professor is to offer knowledge and perspective. As a student, your responsibility is to absorb the useful, discard the useless and build a system for yourself. I can see how a teacher can influence his students (OK, maybe not MBAs). My cousin in Shanghai used to hate English because her English teacher was not nice to her. Isn't that crazy?

# B The Motives for Option Trading

The motives behind options trading could vary from speculation to hedging. Investors with private (legal or illegal) information might choose to trade in the options market to take advantage of the inherent leverage in options. This usually happens more at the level of options on individual stocks, where option investors trade their private information about the idiosyncratic component of the stocks. I have a paper with Allen Poteshman on this topic.

As a graduate from Chicago GSB, Allen was able to get a very unique dataset from CBOE with details on option trading volumes on open buy and sell, close buy and sell from

1990 through 2001. Around the same time, I was teaching 15.433 and had to educate myself about quant investing and sorting portfolios with signals. Like some of you, after learning about this cross-sectional approach, I started to think about trading strategies. Since I spent most of my time thinking about options, the idea came quite naturally to me: would it be cool to have a signal from the options market and use it to trade in the stock market? The most obvious signal would be put/call ratio. Consider a stock with a lot of put option volume traded on it versus a stock with a lot of call option volume traded on it. One is a bearish signal on the stock and the other bullish.

My problem was that I did not have good options data with clean volume information to test this idea. Most of the publicly available data mixes open buy with close buy and open sell with close sell. As a result, the pure signal from open buy is contaminated by close buy. Likewise for the sell volume. So my test results using the publicly available data were weak and I did not want to write a paper with these weak results. This is how I located Allen and his unique dataset. I sent him an email, he sent me a disc with his data and we started to work together.

We form stock portfolios by their put/call ratios and track their performance for the next week. We find that stocks with low put/call ratio outperform stocks with high put/call ratio by 40 basis points over the next day and 1% over the next week. This predictability is stronger for smaller stocks. We also find that option volumes by customers from full service brokerage firms (e.g., hedge funds) are by far the most informative. By contrast, option volumes by firm proprietary traders do not have any predictive power. Our interpretation is that prop traders use exchange-traded options mostly for hedging needs, which is supported by the fact that prop traders are much more active on index options than equity options.

After we finished our paper in 2003 or 2004, we got a lot of interest from practitioners. We even heard from CBOE, who asked us where we got the data. We told them that the data was sitting in their mainframe and offered to help them package and sell the data (so that we can have free access). They said "No." Later they started to sell this data at a pretty high price. Several years later around 2009 or 2010, a former student of Allen got this big grant for data purchase. So I asked her to buy the very expensive CBOE data to do the same test on the more recent data. The strong predictability we found over the 1990-2001 sample no longer exists in the recent time period. It is difficult to say if our paper has any direct impact, but the market seemed to become a little more efficient. After writing this paper, Allen became more interested in the practice of Finance and left his tenured professorship to join D.E. Shaw. He also got us a coverage on the New York Times.

Jun Pan
junpan@mit.edu

253-3083
E62-624

## Classes 14 & 15: Options, Part 2

This Version: November 2, 2016

The Black-Scholes option pricing model, along with the arbitrage-free risk-neutral pricing framework, is something of a revolution in Finance. It managed to attract many mathematicians, physicists, and even engineers to Finance. But if the progression stopped right at the level of modeling and pricing, it would have been rather boring: you take the pricing formula, plug in the numbers, and get the price. So things would have been pretty mechanical. Real life is always more interesting than financial models. In this class, let's bring the model to the data and enjoy the discovery process.

# 1 Bring the Black-Scholes Model to the Data

- **ATM Options and Time-Varying Volatility:** Volatility plays a central role in option pricing. In the Black-Scholes model, volatility $\sigma$ is a constant. If you take this assumption literally, then the Black-Scholes implied vol $\sigma_t^I$ should be a constant over time. In practice, this is not at all true. As we learned in our earlier class on time-varying volatility, using either SMA or EWMA models, the volatility measured from the underlying stock market moves over time. Recall this plot, Figure 1, in Classes 8 & 9, where the option-implied volatility is plotted against the volatility measured directly from the underlying stock market. In both cases, stock return volatility varies over time.

  One interesting observation offered by Figure 1 is that the option-implied volatility is usually higher than the actual realized volatility in the stock market. In other words, within the Black-Scholes model, the options are more expensive than what can be justified by the underlying stock market volatility. If you believe in the Black-Scholes model, then selling volatility (via selling near-the-money options, calls or puts) will be a very profitable trading strategy.

  Figure 2 plots the time-series of VIX (option-implied volatility using SPX) against the time-series of the S&P 500 index level. As you can see, the random shocks to

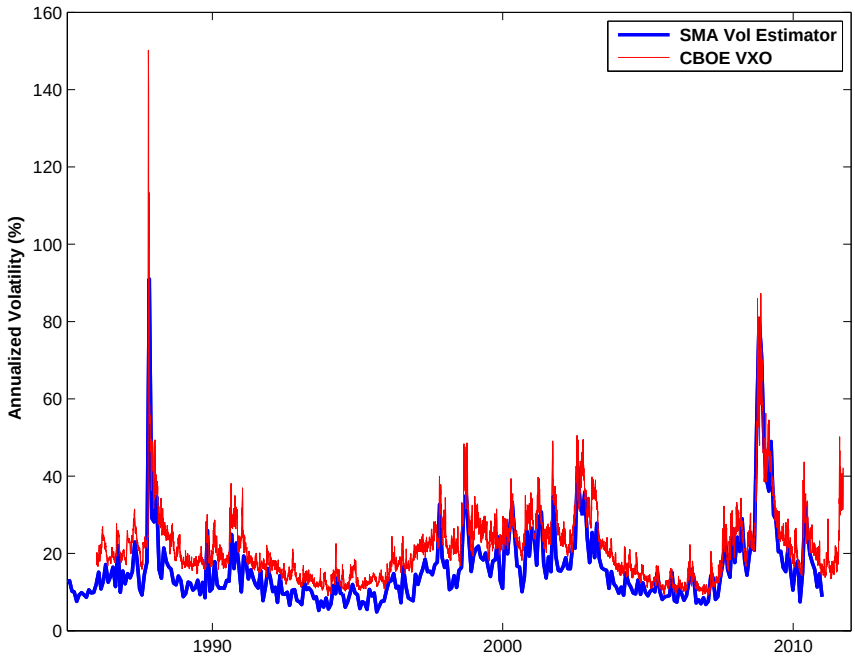


Figure 1: Time-Varying Volatility of the S&P 500 Index. The red line is the option-implied volatility using SPX traded on CBOE. The blue line is measured directly from the underlying stock market using daily returns of the S&P index.

VIX, especially those sudden increases in VIX are often accompanied by sudden and large drops in the index level. Of course, this observation is outside of the Black-Scholes model, where $\sigma$ is a constant. But this plot gives us the intuition as to what could go wrong with selling volatility: you lose money when the markets are in crisis. Basically, by selling volatility on the overall market (e.g., SPX), your capital is at risk exactly when capital is scarce. In the language of the CAPM, you have a positive beta exposure.

But this positive beta exposure is more subtle than the simple linear co-movement captured by beta. As highlighted by the shaded areas, volatility typically spikes up when there are large crises. Just to name a few: the October 1987 stock market crash, the January 1991 Iraq war, the September 1993 Sterling crisis, the 1997 Asian crisis, the 1998 LTCM crisis, 9/11, 2002 Internet bubble burst, 2005 downgrade of GM and Ford, 2007 pre-crisis, March 2008 Bear Stearns, September 2008 Lehman, the European and Greek crises in 2010 and 2011, and the August 2015 Chinese spillover. In other words, what captured by Figure 2 is co-movement in extreme events, like the crisis beta in Assignment 1 (risk exposure conditioning on large negative stock returns). Also, as shown in Figure 2, not all crises have the same impact. For example, the downgrade of GM and Ford was a big event for the credit market, but not too scary for equity and index options.

The comovement in Figure 2 gives rise to a negative correlation between the S&P 500 index returns and changes in VIX, which ranges between -50% to -90%. Figure 3 is an old plot from Classes 8 & 9, which uses the EWMA model to estimate the correlation between the two. As you can see from the plot, the correlation has experienced a regime change. During the early sample period, the correlation hovers around -50%, while in more recent period, the correlation has become more severe, hovering around -80%.

All of these observations have direct impact on how options should be priced in practice: the Black-Scholes model need to allow $\sigma$ to vary over time. The time variation of $\sigma$ should not be modeled in a deterministic fashion. As shown in Figure 2, the time series of $\sigma_t$ is affected by uncertain, random shocks. So just like the stock price $S_t$ follows a stochastic process (e.g., geometric Brownian motion), $\sigma_t$ itself should follow a stochastic process with its own random shocks. Moreover, the random shocks in $\sigma_t$ should be negatively correlated with the random shocks in $S_t$ to match the empirical evidence in Figure 2. There is a class of diffusion models called stochastic volatility models developed exactly for this purpose.

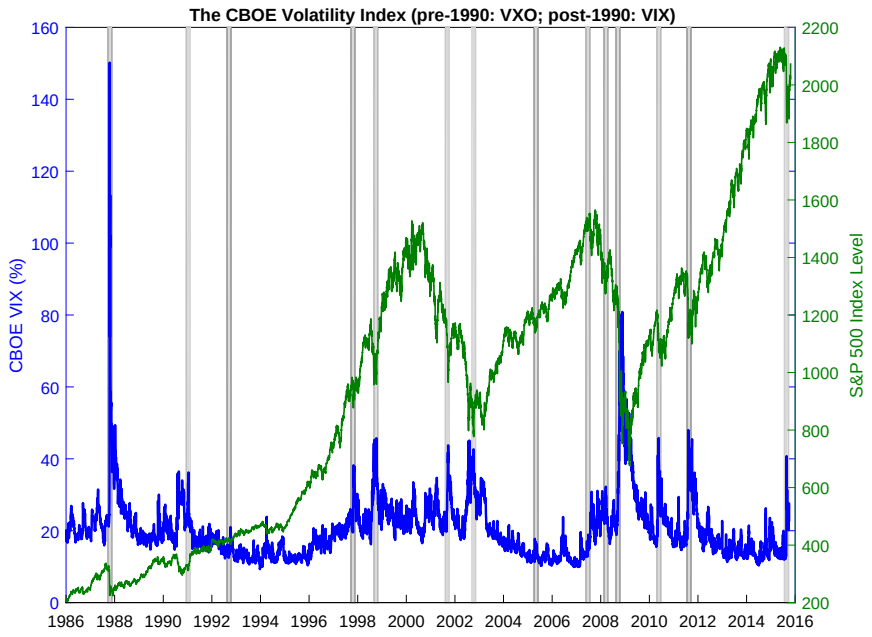


Figure 2: Time-Series of the CBOE VIX Index Plotted against the Time-Series of the S&P 500 Index Level. Prior to 1990, the old VIX (VXO) is used. Post 1990, the news VIX index is used.

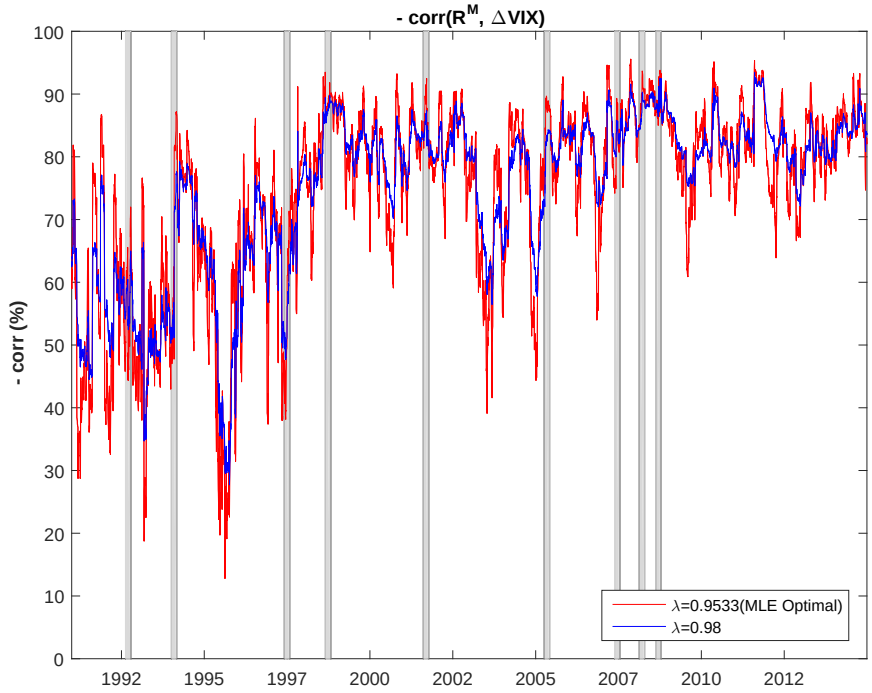


Figure 3: The Time-Series of EWMA Estimates for the Correlations between the S&P 500 Index Returns and Daily Changes in the VIX Index.

These models are similar to the discrete-time models like EWMA or GARCH, which also allow volatility to be time-varying. But one distinct future of stochastic volatility models is that it has its own random shocks. In EWMA or GARCH, the time-varying volatility comes from the random shocks in the stock market. We will come back to the stochastic volatility model later in the class, which are very useful in pricing options of different times to expiration, linking the pricing of long-dated options to that of short-dated options.

- **OTM Options and Tail Events:** In developing our intuition for the Black-Scholes model, we've focused mostly on the ATM options, which are important vehicles for volatility exposure. Now let's look at the pricing of the out-of-the-money options.

  Recall the risk-neutral pricing of a call option,

$$C_0 = E^Q\left(e^{-rT}(S_T - K)\mathbf{1}_{S_T>K}\right) = \boxed{e^{-rT}\,E^Q\left(S_T\mathbf{1}_{S_T>K}\right)} - \boxed{e^{-rT}\,K\,E^Q\left(\mathbf{1}_{S_T>K}\right)},$$

  where the pricing bolts down to calculations involving $E^Q(\mathbf{1}_{S_T>K}$ $E^Q(S_T\mathbf{1}_{S_T>K}$. For $K > S_0e^{rT}$, the call option is out of the money. In fact, the larger the strike price $K$, the more out of the money the option is, and the smaller $E^Q(\mathbf{1}_{S_T>K}$. So if we focus on OTM calls, we zoom into the right tail.

  Likewise, the risk-neutral pricing of a put option is,

$$P_0 = E^Q\left(e^{-rT}(K - S_T)\mathbf{1}_{S_T<K}\right) = \boxed{e^{-rT}\,K\,E^Q\left(\mathbf{1}_{S_T<K}\right)} - \boxed{e^{-rT}\,E^Q\left(S_T\mathbf{1}_{S_T<K}\right)},$$

  where the pricing bolts down to calculations involving $E^Q(\mathbf{1}_{S_T<K}$ $E^Q(S_T\mathbf{1}_{S_T<K}$. For $K < S_0e^{rT}$, the put option is out of the money. In fact, the smaller the strike price $K$, the more out of the money the option is, and the smaller $E^Q(\mathbf{1}_{S_T<K}$. So if we focus on OTM puts, we zoom into the left tail.

  Within the Black-Scholes model, the above calculations can be taken to the next level using the probability distribution of a standard normal:

$$P_0 = \boxed{e^{-rT}\,K\,E^Q\left(\mathbf{1}_{S_T<K}\right)} - \boxed{e^{-rT}\,E^Q\left(S_T\mathbf{1}_{S_T<K}\right)} = \boxed{e^{-rT}\,K\,N(-d_2)} - \boxed{S_0\,N(-d_1)},$$

  where I've changed the color coding so that this equation matches with Figure 4. More specifically, for a 10% OTM put striking at $K = S_0\,e^{r\,T} \times 90\%$, we can re-write the

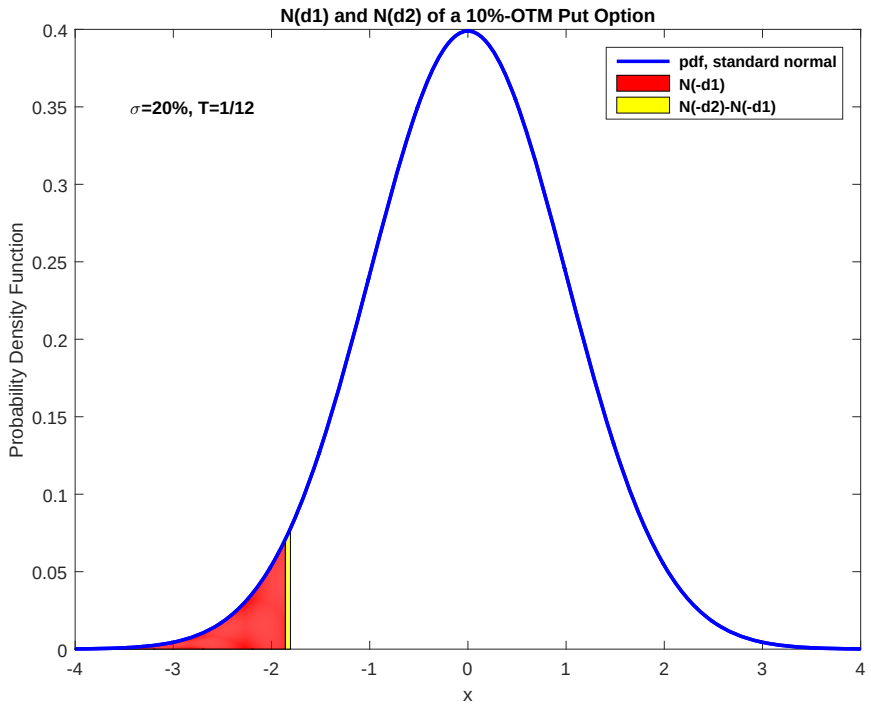


Figure 4: The Distribution of a Standard Normal with $N(-d_1)$ and $N(-d_2)$.

.

above pricing into:

$$\frac{P_0}{S_0} = \frac{e^{-rT}K}{S_0}\boxed{N(-d_2)} - \boxed{N(-d_1)} = 0.90 \times \boxed{N(-d_2)} - \boxed{N(-d_1)}$$

Figure 4 gives us a graphical presentation of what matters when it comes to pricing such OTM options: the left tail in red and the slice in yellow. The areas in red and yellow are mapped directly to the CDF of a standard normal (hence $N(-d_1)$ and $N(-d_2)$ - $N(-d_1)$) because we are working under the Black-Scholes model. But the intuitive goes further. For any distribution (even it is not normal), what matters for the pricing of this OTM put option is the left-tail distribution. If this left tail is fat because of many financial crises, then the pricing of OTM put options should reflect these tail events. In Assignment 3, you will have a chance to work with a model with crash and see the link between fat tails and option prices.

As mentioned a few times, the actual distribution of stock market returns is not normally distributed. This is especially true for returns at higher frequencies (e.g., daily returns). As such, the Black-Scholes model fails to capture the fat tails in the data. As we will see, this becomes a rather important issue when it comes to pricing options.

Conversely, by looking at how these OTM options are priced, we learn about investors' assessment and attitude toward these tail events.

- **Option Implied Smirks:** After the 1987 stock market crash, one very robust pattern arose from the index options (SPX) market called volatility smiles or smirks.

  Consider the nearest term options, say one month to expiration (T=1/12). Let's vary the strike price of these options. Typically, for options with one month to expiration, you can find tradings of OTM puts and calls that are up to 10% out of the money. It is generally the case that OTM options are more actively traded than in-the-money options. This makes sense. If you are using options for speculations, you would prefer options that are cheaper (and are liquid) so that you can get more action for each dollar invested in options. If you are using options for hedging, it is likely that you are hedging out tails events. So either way, the OTM puts and calls are referred instruments than ITM options.

  Between OTM calls and puts of SPX, it is generally the case that OTM puts are more actively traded and the level of OTM-ness can reach up to 20%. For the S&P 500 index, a typical annual volatility is 20%, implying a monthly volatility of 5.77%. So for a 10% OTM put option, it takes a drop of 1.733-sigma (10%/5.77%) move in the S&P 500 index over a one-month period for this option to come back to the money.

  Using the market prices of all the available SPX puts and calls, we can back out the Black-Scholes implied volatility $\sigma^I$ for each one of them. If investors are pricing the options according to the Black-Scholes model, then we should see $\sigma^I$ being exactly the same for all of these options, regardless of the moneyness of the options. What we see in practice, however, is a pattern like that in Table 1.

Table 1: Short-Dated SPX Puts with Varying Moneyness on March 2, 2006.

| $P_0$ | $S_0$ | $K$ | OTM-ness | $T$ | $\sigma^I$ | $P_0^{\text{BS}}$ |
|---|---|---|---|---|---|---|
| 9.30 | 1287 | 1285 | 0.15% | 16/365 | 10.06% | ? |
| 6.00 | 1287 | 1275 | 0.93% | 16/365 | 10.64% | 5.44 |
| 2.20 | 1287 | 1250 | 2.87% | 16/365 | 12.74% | 0.92 |
| 1.20 | 1287 | 1225 | 4.82% | 16/365 | 15.91% | 0.075 |
| 1.00 | 1287 | 1215 | 5.59% | 16/365 | 17.24% | 0.022 |
| 0.40 | 1287 | 1170 | 9.09% | 16/365 | 22.19% | 0.000013 |

Table 1 lists six short-dated OTM put options with exactly the same time to expiration but varying degrees of moneyness. The first option is nearest to the money, striking

at $K = 1285$ when the underlying stock index is at $S_0 = 1287$. The last option is the farthest away from the money, striking at $K = 1170$. The S&P 500 index needs to drop by over 9% over the next 16 calendar days in order for this option to be in the money. Not surprisingly, options are cheaper as they are farther out of the money. But what's interesting is that their Black-Scholes implied vols exhibit this opposite pattern: the more out of the money a put option is, the higher its implied vol. In other words, even though the pricing of \$0.40 (per option on one underlying share of the S&P 500 index) seems very cheap in dollars and cents, it is actually over priced. Plugging a $\sigma = 10.06\%$ to the Black-Scholes model (which a closer to the market volatility around March 2, 2006), the model price for this OTM put is \$0.000013. In other words, this option is so out of the money, the Black-Scholes model (with normal distribution) deems its value to be close to zero. In practice, however, there are people who are willing to pay \$0.40 for it.

Why? Don't they know about the Black-Scholes option pricing formula? If they care about tail events, then what about OTM calls which are sensitive to right tails? As we see in the data, the tail fatness shows up in both the left and the right. But the OTM calls are not over-priced. If anything, the implied vols of OTM calls are on average slightly lower than ATM options. That is why we are calling this pattern volatility smirk, which is an asymmetric smile.

- **Expected Option Returns:** Another way to look at the profit/loss involved in options is to calculation their expected returns like we do in the stock market. Table 2 was reported in a 2000 *Journal of Finance* paper by Prof. Coval and Shumway.

Table 2: Expected Options Returns

| Strike - Spot | -15 to -10 | -10 to -5 | -5 to 0 | 0 to 5 | 5 to 10 |
|---|---|---|---|---|---|
| Weekly SPX Put Option Returns (in %) | | | | | |
| mean return | -14.56 | -12.78 | -9.50 | -7.71 | -6.16 |
| max return | 475.88 | 359.18 | 307.88 | 228.57 | 174.70 |
| min return | -84.03 | -84.72 | -87.72 | -88.90 | -85.98 |
| mean BS $\beta$ | -36.85 | -37.53 | -35.23 | -31.11 | -26.53 |
| corrected return | -10.31 | -8.45 | -5.44 | -4.12 | -3.10 |

Option data from Jan. 1990 through Oct. 1995.

As shown in Table 2, the weekly returns of buying put options are on average negative. There are quite a bit of variation in these returns. For the farther OTM put options,

the return could be as positive as 475.88%, or as negative as -84.03%. This option has a beta of -36.85, which is due to the inherent leverage of these options. The CAPM-alpha of this investment is -10.31% per week. Whoever is selling this option would make a lot of money...on average. But he needs to be well capitalized when an event like 475.88% happens.

Calculations like those in Table 2 are rather imprecise because of the large variations in option returns. So we do not want to take the numbers too literally. But the qualitative result of this Table is important: when it comes to investing in options, there are large variations in option returns. Moreover, buying put options give you negative alpha. The more out of the money the put option is, the more negative the alpha becomes. For investors who are selling such put options, they are able to capture such alpha. But such trading strategies are in generally very dangerous. You need to be well capitalized to survive large crises like the 1987 stock market crash. Otherwise, you are just one crisis away from bankruptcy.

The results shown here in the return space is very much consistent with the earlier results in the implied-vol space, where OTM put options are over priced relative to near-the-money options. The level of over-pricing gets more severe as the put option becomes more out of the money and are more sensitive to market crashes. So it is not surprising that the put option returns are on average negative. Most of the times, you purchase an insurance against a market crash, but the crash does not happen and your put option expires out of the money. But once in a while, a crisis like 1987 or 2008 happens, then this put option brings you over-sized returns. Sitting on the other side of the trade are investors who sell/write you these crash insurances. Most of the times, they are able to pocket the premiums paid for the insurance without having to do anything. But once in while, they lose quite a bit of money if a crisis like 1987 or 2008 happens. As such, the risk profile of such option strategies differs quite significantly from that of a stock portfolio, where all instruments are linear. In Assignment 3, you will have a chance to see this kind of risk/return tradeoff of options in more details for yourselves.

# 2 When Crash Happens

- **Crash and Crash Premium:** The empirical evidence we've seen so far indicates that strategies involving selling volatility and selling crash insurance are profitable. As you will see for yourself in Assignment 3, the return distribution of such option strategies

differs quite significantly from that of a stock portfolio, where all instruments are linear.

In the presence of tail risk, options are no longer redundant and cannot be dynamically replicated. As such, two considerations involving the tail risk become important in the pricing of options. First, the likelihood and magnitude of the tail risk. Second, investor's aversion or preferences toward such tail events. The "over-pricing" of put options on the aggregate stock market (e.g., the S&P 500 index) reflects not only the probability and severity of market crashes, but also investors' aversion to such crashes — crash premium.

In fact, as you will see in Assignment 3, the probability and severity of market crashes implicit in the volatility smirk are such that investors are pricing these OTM put options as if crashes like 1987 would happen at a much higher frequency. In other words, investors are willing to pay a higher price for such crash insurances even though they are "over-priced" relative to the actual amount of tail risk observed in the aggregate stock market. And the sellers of such crash insurances are only willing to sell them if they are being compensated with a premium, above and beyond the amount of tail risk in the data. This crash premium accounts for most of the "over-pricing" in short-dated OTM puts and ATM options.

By contrast, this "over-pricing" is not severe for OTM calls because they are not very sensitive to the left tail. Instead, OTM calls are sensitive to the right tail. From how such options are priced relative to OTM puts, it is obvious that investors are not eager to pay the same amount of premium for insurances against the right tail. This makes perfect sense. The intuition comes straight from the CAPM. An OTM call is a positive beta security, which provides positive returns when the market is doing well. It is icing on the cake. By contrast, an OTM put pays when the market is in trouble — a friend in need is a friend indeed.

- **Bank of Volatility:** LTCM was a hedge fund initially specialized in fixed-income arbitrage. It was extremely successful in its earlier years. Success breeds imitation. Soon, the fixed-income arbitrage space was crowded and spreads in arbitrage trades were shrinking. In early 1998, LTCM began to short large amounts of equity volatility. Betting that implied vol would eventually revert to its long-run mean of 15%, they shorted options at prices with an implied volatility of 19%. Their position is such that each percentage change in implied vol will make or lose $40 million in their option portfolio. The size of their vol position was so big that Morgan Stanley coined a nickname for the fund: the Central Bank of Volatility. For more details, you can read

Roger Lowenstein's book on LTCM.

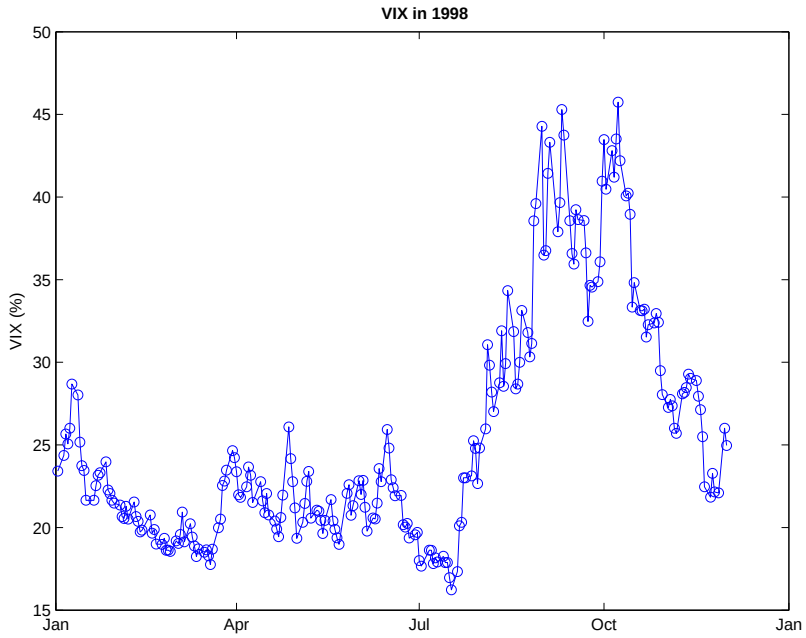


Figure 5: Time Series of CBOE VIX index in 1998.

During normal time, volatility does revert to its mean. So the idea behind the trade makes sense. Moreover, as we've seen in the data, selling volatility (via ATM options) is a profitable strategy on average because of the premium component. But the premium was not a free lunch: it exists become of the risk involved in selling volatility. As we've seen in the data, when the volatility of the aggregate market suddenly spikes up, the financial market usually is in trouble. Whenever the market is in the crisis mode, there is flight to quality: investor abandon all risky asset classes and move their capital to safe havens such as the Treasury bond market.

For the case of LTCM in 1998, it had arbitrage trades in different markets (e.g., equity, fixed-income, credit, currency, and derivatives) across different geographical locations (e.g., U.S., Japan, and European). Lowenstein's book gives more detailed descriptions of these arbitrage trades. One common characteristics of these arbitrage trades is that they locate some temporary dislocation in the market and speculate that this dislocation will die out as the market converges back to normal. In a way, these arbitrage trades betting on convergence make money because they provide liquidity to temporary market dislocations. The key risk involved in these arbitrage trades is that timing of the convergence is uncertain. Sometimes, instead of converging, the

dislocation becomes even more severe before converging back to normal.

Prior to the Russian default in the summer of 1998, these arbitrage trades were not highly correlated. But after the default, most of these previously uncorrelated arbitrage trades lost money for LTCM at the same time. This certainly includes the volatility trades. As shown in Figure 5, early in the year, volatility was fluctuating around 20%. By summer 1998, however, the market became quite volatile because of the Russian default. At its peak, the VIX index was around 45%. Recall that LTCM was selling volatility when VIX was around 19% in early 1998. The position was such that each percentage change in implied vol will make or lose $40 million. So if the volatility converges back to its long run mean of 15%, then roughly $4 \times \$40 = \$160$ can be made. But if instead of converging, the volatility increases to 45%, you can imagine the loss.

The Russian default affected not only LTCM but other hedge funds and prop trading desks who pursued the same kind of convergence trades. At a time like this, capital becomes scarce, and all leveraged investors (e.g., hedge funds or prop trading in investment banks) are desperately looking for extra source of funding. They do so by unwinding some of their arbitrage trades, further exacerbating the widening spreads. At a time like this, holding a security that pays (e.g., an existing long position in puts) could be very valuable. By contrast, a security that demands payment (e.g., an existing short position in puts) would be threatening to your survival. Therefore, being on the short side of the market volatility hurts during crises. That is why volatility is expensive (i.e., ATM options are over-priced) in the first place.

- **The 2008 crisis:** The OTM put options on the S&P 500 index is a good example for us to understand crash insurance. In writing a deep OTM put option, the investor prepares himself for the worse case scenario when the option becomes in the money. This happens when the overall market experiences a sharp decline. The probability of such events is small. But if he writes a lot of such options believing that the exposure can somehow be contained by the low probability, then he is up for a big surprise when a crisis does happen. As we learned from the recent financial crisis, some supposedly sophisticated investors wrote such OTM put options without knowing the real consequence.

  Gillan Tett from *Financial Times* wrote an excellent book called *Fool's Gold* with details of how investment banks developed and later competed for the market shares of the mortgage-linked CDO products. The following is a brief summary.

  By 2006, Merrill, who was late into the CDO game, topped the league table in terms

of underwriting CDO's, selling a total of $52 billion that year, up from $2 billion in 2001. Behind the scenes, Merrill was facing the same problem that worried Winters at J.P. Morgan: what to do with the super-senior tranch?

CDO's are the collateralized debt obligations. It pools individual debt together and slices the pool into tranches according to seniority. For a mortgage-linked CDO, the underlying pool consists of mortgages of individual homeowners. The cashflow to the pool consists of their monthly mortgage payments. The most senior tranch is the first in line to receive this cashflow. Only after the senior tranch receives its promised cashflow, the next level of tranches (often called mezzanine tranches) can claim their promised cashflow. The equity tranch is the most junior and receives the residual cashflow from the pool.

As default increases in mortgages, the cashflow to the pool decreases. The equity investors will be the first to be hit by the default. If the default rate further increases, then the mezzanine tranch will be affected. The most senior tranch will only be affected in the unlikely event that both equity and mezzanine investors are wiped out and the cashflow to the pool cannot meet the promise to the most senior tranch. Such super senior tranches are usually very safe and are Aaa rated. By contrast, the mezzanine tranches are lower rated (Baa) because of the higher default risk. And the credit quality of the equity tranch is even lower.

The pricing of such products is consistent with their credit quality: the yield on the mezzanine tranches is higher than the senior tranches to compensate for the higher credit risk. Investors, in an effort to reach for yield, prefer to buy the mezzanine and equity tranches. As a result, the investment banks underwriting the CDOs are often stuck with the super senior tranches. As the business of CDOs grew, the banks are accumulating more and more highly rated super senior tranches. Initially, Merrill solved the problem by buying insurance (credit default swaps) for its super-senior debt from AIG.

Let's take a look at what the super-senior tranch is really about. It is highly rated because of the low credit risk. Imagine the economic condition under which this credit risk affecting the super-senior tranch will actually materialize: when the default risk is so high that both mezzanine and equity investors are wiped out. A typical argument for the economics of pooling is that default risk by individual homeowners can be diversified in a pool. This is indeed true when we think about the risk affecting the equity tranch: one or two defaults in the pool would affect the cashflow to the equity tranch, but would not affect the mezzanine tranch, let alone the senior tranch. So the

risk affecting the senior tranch has to be a very severe one. The default rate has to be so high that the cashflow dwindles to the extent that it would eat through the lower tranches and affect the most senior tranch. In other words, many homeowners must be affected simultaneously and default at the same time to generate this type of scenario. By then, the risk is no longer idiosyncratic but systemic. So writing an insurance on a senior tranch amounts to insuring a crisis — a deep OTM put option on the entire economy.

In late 2005, AIG told Merrill that it would no longer offer the service of writing insurance on senior tranches. By then, however, AIG has already accumulated quite a large position on such insurance. Later, AIG was taken over by the US government in a $85 billion bailout and the insurance on senior tranches was honored and made whole by AIG (and the New York Fed).

After AIG declined to insure their super senior tranch, Merrill decided to start keeping the risk on its own books. At the same time, Citigroup, another late comer, was also keen to ramp up the output of its CDO machine. Unlike the brokerages, though, Citi could not park unlimited quantities of super-senior tranches on its balance sheet. Citi decided to circumvent that rule by placing large volumes of its super-senior in an extensive network of SIVs (Special Investment Vehicle) and other off balance sheet vehicles that it created. Citi further promised to buy bank the super-senior tranch if the SIVs ever ran into problems with them.

Now let's try to understand what Merrill and Citi are actually doing by retaining the super-senior tranch. Effectively, they are holding the super-senior tranch without an insurance. If you are holding a US treasury bond, you don't have to worry about credit risk (except for when the US government defaulted). So holding a super-senior tranch without an insurance is like holding a default-free US treasury bond and selling a deep OTM option on the overall economy at the same time. Before, they were able to buy that put option from AIG to hedge out this risk. Now, they are bearing this risk themselves.

Then the crisis happened in 2007 and 2008, and the mortgage default rate increased to such an extent that it started to affect the super-senior tranches. In other words, the deep OTM put options became in the money. During the 2007-08 crisis, the pricing of these super-senior tranches became one of the biggest headaches on Wall Street. Merrill and Citi, along with other Wall Street banks, had to take billions of dollars of writedowns.

# 3 Beyond the Black-Scholes Model

- **A model with market crash:** In Assignment 3, you will be working closely with a model that allows market to crash. It is a simplifies version of the model in Merton (1976).
- **A model with stochastic volatility:** I'll briefly mention these models in class.

# Appendix

During my office hours, I got a few questions about the Brownian motion and risk-neutral pricing. Let me use this appendix to explain some of the details.

## A Brownian Motion

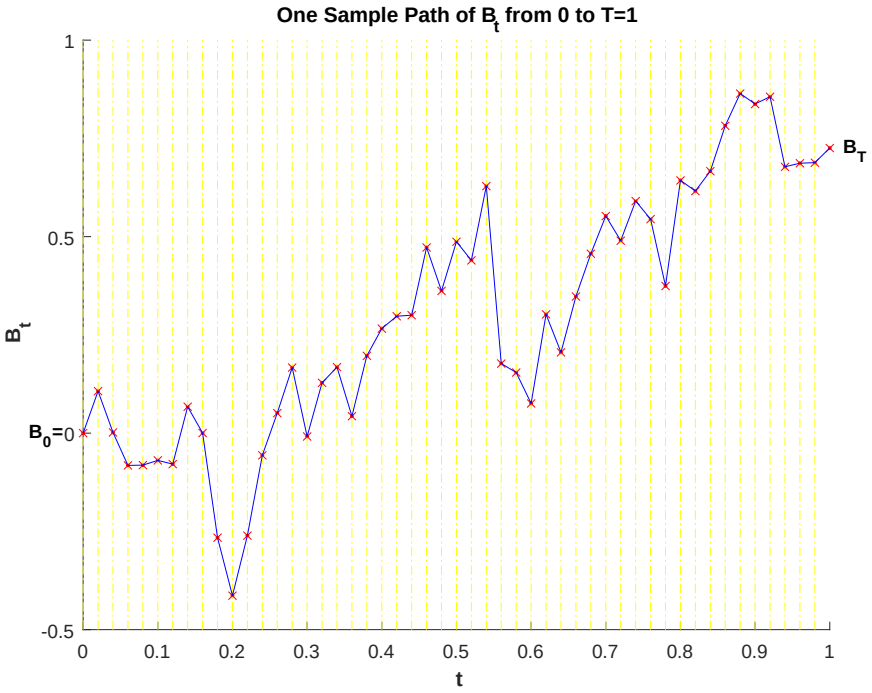


Figure 6: One sample path of a Brownian motion.

To understand the Brownian motion, let's create one. Let's start from time 0 and end in time T. Let's further chop this time interval into small increments. For example, in Figure 6, T=1 and the interval between 0 and 1 is chopped evenly into 50 smaller increments with size $\Delta = 1/50$. We can now start to create a sample path of the Brownian motion:

$$
\begin{aligned}
&B_0 = 0 \\
&B_\Delta - B_0 = \sqrt{\Delta}\,\epsilon_\Delta \\
&B_{2\Delta} - B_\Delta = \sqrt{\Delta}\,\epsilon_{2\Delta} \\
&\dots \\
&B_T - B_{T-\Delta} = \sqrt{\Delta}\,\epsilon_T\,.
\end{aligned}
$$

where the $\epsilon$'s are independent standard normals. In creating this sample path, we use the first two properties of the Brownian motions: independence increments and stationary normal increments. I've attached the Matlab code I used to create this plot in this note. You can run it and each time you will get a different sample path.

For our purpose of pricing a European-style option, what matters is the distribution of $B_T$. But if we are interested in pricing an American-style option, then the entire path of $B_t$ matters and at each node, we will make a decision of whether or not to exercise early. So the grid should be as fine as possible (larger $N$ and smaller $\Delta$).

Now back to our original process for $S_t$:

$$dS_t = \mu\, S_t\, dt + \sigma\, S_t\, dB_t\,,$$

where to avoid distraction, I have set the dividend yield $q = 0$. As usual, we work with $X_t = \ln S_t$ and, using the Ito's Lemma, we have

$$dX_t = \left(\mu - \frac{1}{2}\sigma^2\right)\, dt + \sigma\, dB_t\,.$$

The nice thing about working with $\ln S_t$ is that you can integrate out the process:

$$\begin{aligned} X_T &= X_0 + \int_0^T \left(\mu - \frac{1}{2}\sigma^2\right)\, dt + \int_0^T \sigma\, dB_t \\ &= X_0 + \left(\mu - \frac{1}{2}\sigma^2\right)\, T + \sigma\, (B_T - B_0) \\ &= X_0 + \left(\mu - \frac{1}{2}\sigma^2\right)\, T + \sigma\, \sqrt{T}\, \epsilon_T\,, \end{aligned}$$

where in the last step I use $\sqrt{T}\,\epsilon_T$ to express $B_T - B_0$. Recall that the log-return $R_T$ is defined by $R_T = \ln X_T - \ln X_0$. We have

$$R_T = \left(\mu - \frac{1}{2}\sigma^2\right)\, T + \sigma\, \sqrt{T}\, \epsilon_T$$

# B Change of Measure, Risk-Neutral Pricing

Under the original measure (P-measure), the process runs as

$$dX_t = \left(\mu - \frac{1}{2}\sigma^2\right)\, dt + \sigma\, dB_t\,.$$

If we were to do option pricing under this measure, we know that we cannot do

$$C_0 \neq e^{-rT} E\left(S_T - K\right)^+ .$$

This is a big no-no in Finance because it approaches the pricing as if we were risk neutral. Interestingly, this is why the method of "risk-neutral" pricing arises. It is mostly a mathematical result. If you Google Girsonov theorem or the Radon-Nikodym derivative, you will see the related math result. But the math result has its relevance in Finance. Let me approach it this way.

In Finance, we develop this concept of pricing kernel or the stochastic discount factor. Armed with this pricing kernel $\xi_T$, we can do our pricing:

$$C_0 = e^{-rT} E\left(\frac{\xi_T}{\xi_0}\left(S_T - K\right)^+\right) .$$

Under the Black-Scholes setting, the markets are complete and the pricing kernel is unique. In fact, as an application of the Girsonov theorem, this pricing kernel is of the form

$$\xi_T = \frac{dQ}{dP} = e^{-\gamma B_T - \frac{1}{2}\gamma^2 T} .$$

This $\xi_T$ is what the mathematician would call the Radon-Nikodym derivative. Notice that by construction $E(\xi_T) = 1$.

As mentioned earlier, the pricing kernel is unique under the Black-Scholes setting. So the constant $\gamma$ is uniquely defined. In Finance, we call this parameter the market price of risk and for the Black-Scholes setting, it is $\gamma = (\mu - r)/\sigma$, which in fact is the Sharpe ratio. In a more general setting, $\gamma$ can itself be a stochastic process. Also notice that with a positive market price of risk, $\gamma > 0$, $\xi_T$ is negatively correlated with $B_T$ (hence negatively correlated with $X_T$ and $S_T$). This is what you were taught in Finance 15.415. When $S_T$ experiences a positive stock, the stochastic discount factor is smaller; when $S_T$ experiences a negative stock, the stochastic discount factor is bigger. This asymmetry has its origin in the fact that investors are risk averse and the risk in $S_T$ is systematic (undiversifiable).

It turns out that we can create a new measure Q, called the equivalent martingale measure, for the original $P$ and the pricing becomes,

$$\begin{aligned} C_0 &= e^{-rT} E\left(\frac{\xi_T}{\xi_0}\left(S_T - K\right)^+\right) \\ &= e^{-rT} E^Q\left(\left(S_T - K\right)^+\right) , \end{aligned}$$

and the link between these two measures is $\xi_T = dQ/dP$.

Now let's construct this new $Q$-Brownian:

$$\begin{aligned} dX_t &= \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma\, dB_t^p \\ &= \left(r - \frac{1}{2}\sigma^2\right) dt + \sigma \left(\frac{\mu - r}{\sigma} + dB_t^P\right) \\ &= \left(r - \frac{1}{2}\sigma^2\right) dt + \sigma\, dB_t^Q\,, \end{aligned}$$

where the Q-Brownian is defined as

$$dB_t^Q = \frac{\mu - r}{\sigma} + dB_t^P\,.$$

And this change of measure, from P to Q, is the essence of the risk-neutral pricing.

The name of "risk-neutral" pricing is ironical: the whole thing arises from the observation that we cannot do

$$C_0 \neq e^{-rT} E^P \left(S_T - K\right)^+\,.$$

But if we are willing to change our probability measure from P to Q, under which

$$dX_t = \left(r - \frac{1}{2}\sigma^2\right) dt + \sigma dB_t^Q\,,$$

then we can indeed do

$$C_0 = e^{-rT} E^Q \left(S_T - K\right)^+\,.$$

# C Change of Measure, One More Application

This mathematical tool can be further exploited. Recall that we need to do this calculation in our Black-Scholes option pricing,

$$e^{-rT}\, E^Q \left(S_T \mathbf{1}_{S_T > K}\right)$$

What if we can drop $S_T$ and change it to

$$S_0\, E^{?} \left(\mathbf{1}_{S_T > K}\right)$$

That would make our math very simple.

In fact, we can drop $S_T$ like the way we dropped $\xi_T$. As long as the process is positive, there is an equivalent martingle measure waiting for us to help us simplify the math. This is where the new measure $QQ$ comes from. You can start with the observation that

$$S_T = e^{X_T} = e^{\sigma B_T + \text{other deterministic terms}}$$

You can then check

$$\begin{aligned} dX_t &= \left(r - \frac{1}{2}\sigma^2\right) dt + \sigma dB_t^Q \\ &= \left(r + \frac{1}{2}\sigma^2\right) dt - \sigma^2 dt + \sigma dB_t^Q \\ &= \left(r + \frac{1}{2}\sigma^2\right) dt + \sigma \left(-\sigma dt + dB_t^Q\right) \end{aligned}$$

So if we define

$$dB_t^{QQ} = -\sigma dt + dB_t^Q ,$$

under which

$$dX_t = \left(r + \frac{1}{2}\sigma^2\right) dt + \sigma dB_t^{QQ} .$$

Then we can indeed get

$$e^{-rT} E^Q \left(S_T \mathbf{1}_{S_T > K}\right) = S_0 E^{QQ} \left(\mathbf{1}_{S_T > K}\right) .$$

I am being a bit sloppy in my notation, but I trust a careful and thorough student would fill in the details (including the *other deterministic terms*).

# D Matlab Code

Code 1: Brownian.m

```
T=1; N=50;

Delta=T/N;
EPS=randn(N,1);
T_vec=(0:Delta:T)';

B=0; B_vec=B;
```

```
for i=1:N,
  B=B+EPS(i)*sqrt(Delta);
  B_vec=[B_vec; B];
end

figure(1); clf; hold on;
plot(T_vec,B_vec,'r.');
BND=axis;
for i=1:N,
 plot(T_vec(i)*[1 1],BND(3:4),'y-.');
end
plot(T_vec,B_vec,'rx',T_vec,B_vec,'b-');
hold off;
ylabel('\bf B_t');
xlabel('\bf t');
text(1.01,B,'\bf B_T');
text(-0.08,B_vec(1),'\bf B_0=');
title('One Sample Path of B_t from 0 to T=1')
```

Jun Pan
junpan@mit.edu

253-3083
E62-624

**Classes 17 & 18: Risk Management**

This Version: November 16, 2016

# 1 Why Risk Management?

- **Capital Markets Imperfection:** According to Modigliani and Miller (1958), in perfect capital markets, adding or subtracting financial risk has no impact on the market value of a publicly traded corporation or on the welfare of its shareholders. In the real world, capital markets are imperfect. This imperfection gives rise to the need for risk management.

  At the core of risk management for financial institutions is the concept of "capital adequacy." If new capital could be obtained in perfect financial markets, we would expect a financial firm to raise capital as necessary to avoid the cost of financial distress. In such a setting, purely financial risk would have a relatively small impact, and risk management would not be important. In practice, however, capital is a scarce resource, especially when it is most needed.

  Compared with other types of corporations, financial firms have relatively more liquid balance sheets, made up largely of financial assets. This relative liquidity allows a typical financial firm to operate with a high degree of leverage. For example, major broker-dealers regulated by SEC frequently have a level of accounting capital that is close to the regulatory minimum of 8% of accounting assets, implying a leverage ratio on the order of 12.5 to 1. As we will see later in the class, for Goldman Sachs, the ratio of book assets to book equity was 10.3 to 1 in 2014, with 23% of the liabilities financed by long-term liabilities and 11% financed by Repo (usually overnight and short-term). In 2007, the leverage was even higher: the asset-to-equity ratio at 26.2; and the financing leaned more toward short term: only 18% long-term financing and close to 15% Repo financing.

  Ironically, in light of the relatively high degree of liquidity that fosters high leverage, a significant and sudden financial loss (or reduced access to credit) can cause dramatic

illiquidity effects. This, has been the experience for many financial firms during the 2007-08 crisis. Some survived (e.g., Morgan Stanley and Goldman Sachs), some were bought out (e.g., Bear Stearns, Merrill, and Wachovia), and some failed (e.g., Lehman and WaMu). For individual firms, weathering sudden financial losses with adequate capital matters for its own survival. For regulators, it is about the financial stability of the entire system, which has become highly inter-connected through interbank transactions including OTC derivatives trading.

- **Liquidity Mismatch in Assets and Liabilities:** Let's strip the complexity of a financial institution to its bare minimum with this simple example of a bank. It takes in deposits at the short-end of the yield curve and makes loans at the long-end. This maturity transformation is at the core of a bank's profitability. As we will learn in the fixed-income class, the yield curve is typically upward sloping with the spread between the long- and short-term yields averaged to about 100 to 200 basis points. In addition, the longer maturity loans made by banks to firms are usually defaultable, adding another 100 to 200 basis points of credit spread (assuming the loans are investment grade).

  With fractional-reserve banking, the bank is allowed to hold reserves that are only a fraction (e.g., 10%) of their deposit liabilities. For our example, let's assume that the bank is 100% financed by liabilities. It takes in 100 dollars of demand deposits (i.e., liabilities), holds 10 dollars of reserve (i.e., cash or safe assets) and lends out 90 dollars in longer maturity and defaultable loans (i.e., risky assets).

  A run on a bank happens when depositors suspect that the bank has made bad investments in its risky loans and is no longer solvent. They rush to the bank simultaneously to withdraw their deposits. While the demand deposits are highly liquid and can be withdrawn in a moment's notice, the loans sitting on the asset side of the bank's balance sheet are typically of longer maturity and not as liquid. This liquidity mismatch between a bank's assets and liabilities is the root cause of a bank run: an otherwise solvent bank needs to raise capital quickly to meet the simultaneous demands from panicking depositors acting out of fear.

  In a perfect financial market, the bank should be able to raise additional funding using its loans as collateral. But because of information asymmetry regarding the credit worthiness of the loans, potential investors are reluctant to extend funding to the bank (with such a short notice and under a bank run scenario). Moreover, if this bank run happens during a crisis, then capital is even more scarce, making it more difficult for

the bank to raise new funding.

So the most likely action of the bank is to sell its long-term assets, often hastily and at fire-sale prices. If multiple banks are facing runs at the same time during a crisis situation, then they would be selling similar long-term assets at severely discounted fire-sale prices. In the U.S., the FDIC deposit insurance has been an effective way to stem out bank runs of this kind. Knowing that their deposits are safely guaranteed by FDIC (up to a certain dollar amount for each depositor), depositors will not rush to the bank to withdraw simultaneously. Consequently, bank runs purely due to liquidity mismatch can be avoided.

- **Equity as a Buffer:** I made the previous example as simple as possible so that we can focus on the heart of the issue: liquidity mismatch. To make the example more realistic, we can further add an equity piece. In doing so, we learn another very important concept: the role of equity as a buffer for risk management.

  Suppose the bank is now financed by 90% liabilities (i.e., demand deposits) and 10% equity. Let's keep the same allocation between risky and riskless assets: 90% risky loans and 10% cash. Now let's see how the 10% equity piece can function as a buffer to cushion the fall of the bank. Suppose the bank has already experienced deposit withdrawal of 10 dollars and has exhausted its 10 dollars of reserves. As the next dollar of withdraw comes in, the bank has to sell a piece of its risky loans. Suppose the fire sale price is 50% of the initial value. To raise 1 dollar of cash, the bank therefore has to sell 2 dollars (book value) of risky loans, incurring a one-dollar loss due to the fire sale. Now the total book value of assets are 88 dollars (90-2), the liabilities are at 79 dollars (90-10-1), and equity absorbs the one-dollar write-down and is at 9 dollars.

  As you can see, equity functions as a buffer to cushion the fall of the bank. Without this equity piece, the bank would have been insolvent. It is obvious that a higher capital ratio (Equity/Asset) adds more buffer and strengthens the financial health of a bank. As you will see later in the class, capital ratios of various kinds are an important part of the regulatory requirements for banks. Although they come in different varieties, depending on how assets and equity are calculated, the essence of these capital ratios is to evaluate the capital adequacy (i.e., the thickness of the buffer) of a bank. In this example, the equity/asset ratio is 10/100=10%. Using the approach of risk-weighted assets (RWA), where cash counts as zero, the RWA of the bank is 90 dollars. Then the capital ratio is 10/90=11.11%. You will find this simple model to be quite handy as we discuss capital ratios for the banking industry.

- **The Balance Sheet of Goldman:** For a financial intermediary such as Goldman Sachs, its balance sheet is certainly more complex than that of a simple bank. But the basic idea is similar.

  So let's start with Goldman's 10K reports. Table 1 summarizes the company's assets, liabilities, and shareholders' equity for a few selected years. You must have learned how to read a financial statement from your accounting classes. So let me focus only on the items that are important for us. I've also changed the names of a few items so that the table would fit in one page.

  Before getting into details, let me mention a few events that are important for Goldman. The company went public in 1999. the first 10K form was published in 1999 with 197 pages. Between 1999 and 2006, the length of the 10K forms fluctuated between 103 pages in 2001 and 298 pages in 2005. The 2007 10K form had 372 pages. On September 21, 2008, Goldman became a bank holding company and the Federal Reserve Board became its primary regulator. Its 2008 form has 731 pages, followed by 411 in 2009, 336 in 2010, 367 in 2011, 480 in 2012, 366 in 2013, and 410 in 2014.

  In conjunction with the increasing thickness of the 10K forms, financial intermediaries like Goldman are facing increasing reporting requirements. Indeed, the financial services industry has been the subject of intense regulatory scrutiny in recent years. The 2010 Dodd-Frank Act significantly altered the financial regulatory regime within which Goldman operates. The implementations of Dodd-Frank and Basel III are still on going, which would have a direct and significant impact on the risk-management practice of this industry.

  - **Assets:** Now let's focus our attention on Table 1. As of December 2014, Goldman holds assets in total of $856 billion. For our purpose, the item that matters the most is "financial instruments owned," which is also the largest item, valued at $312 billion. Going back to our example of a simple bank, this item is similar to the risky loans made by a bank. In the case of Goldman, of course, the collection of risky assets is more diverse. We will focus on this item shortly.

    The two items under "collateralized agreements" are effectively collateralized lending, which are relatively safe in terms of market and credit risk, but are subject to counterparty credit risk. Likewise, items under "receivables" are also sensitive to counterparty credit risk. For the purpose of risk management, measuring and controlling counterparty credit risk is an important component, as you will see later, these items show up in the firm's credit risk weighted assets.

Table 1: Goldman Sachs' Assets, Liabilities, and Shareholders' Equity

| **Assets** | | | | |
|---|---|---|---|---|
| in millions | 2014 | 2010 | 2008 | 2007 |
| Cash and cash equivalents | 57,600 | 39,788 | 15,740 | 10,282 |
| Cash and securities for regulatory and other purposes | 51,716 | 53,731 | 106,664 | 119,939 |
| Collateralized agreements: | | | | |
| Repo Lending and federal funds sold | 127,938 | 188,355 | 122,021 | 87,317 |
| Securities borrowed | 160,722 | 166,306 | 180,795 | 277,413 |
| Receivables: | | | | |
| Brokers, dealers and clearing organizations | 30,671 | 10,437 | 25,899 | 19,078 |
| Customers and counterparties | 63,808 | 67,703 | 64,665 | 129,105 |
| Loans receivable | 28,938 | | | |
| **Financial instruments owned** | 312,248 | 356,953 | 328,325 | 452,595 |
| Other assets | 22,599 | 28,059 | 30,438 | 24,067 |
| **Total assets** | **856,240** | **911,332** | **884,547** | **1,119,796** |

| **Liability and Shareholders' Equity** | | | | |
|---|---|---|---|---|
| in millions | 2014 | 2010 | 2008 | 2007 |
| Deposits | 83,008 | 38,569 | 27,643 | 15,370 |
| Collateralized financings | | | | |
| Repo financing | 88,215 | 162,345 | 62,883 | 159,178 |
| Securities loaned | 5,570 | 11,212 | 17,060 | 28,624 |
| Other | 22,809 | 38,377 | 38,683 | 65,710 |
| Payables: | | | | |
| Brokers, dealers and clearing organizations | 6,636 | 3,234 | 8,585 | 8,335 |
| Customers and counterparties | 206,936 | 187,270 | 245,258 | 310,118 |
| **Financial instruments sold short** | 132,083 | 140,717 | 175,972 | 215,023 |
| Unsecured short-term borrowings | 44,540 | 47,842 | 52,658 | 71,557 |
| Unsecured long-term borrowings | 167,571 | 174,399 | 168,220 | 164,174 |
| Other liabilities and accrued expenses | 16,075 | 30,011 | 23,216 | 38,907 |
| **Total liabilities** | **773,443** | **833,976** | **820,178** | **1,076,996** |
| **Total shareholders' equity** | **82,797** | **77,356** | **64,369** | **42,800** |

| $ in millions | As of December 2014 Financial Instruments Owned | Financial Instruments Sold, But Not Yet Purchased |
|---|---|---|
| Commercial paper, certificates of deposit, time deposits and other money market instruments | $ 3,654 | $ — |
| U.S. government and federal agency obligations | 48,002 | 12,762 |
| Non-U.S. government and agency obligations | 37,059 | 20,500 |
| Mortgage and other asset-backed loans and securities: | | |
| Loans and securities backed by commercial real estate | 6,582[1] | 1 |
| Loans and securities backed by residential real estate | 11,717[2] | — |
| Bank loans and bridge loans | 15,613 | 464[4] |
| Corporate debt securities | 21,603 | 5,800 |
| State and municipal obligations | 1,203 | — |
| Other debt obligations | 3,257[3] | 2 |
| Equities and convertible debentures | 96,442 | 28,314 |
| Commodities | 3,846 | 1,224 |
| Subtotal | 248,978 | 69,067 |
| Derivatives | 63,270 | 63,016 |
| **Total** | **$312,248** | **$132,083** |

Figure 1: Goldman's financial instruments, long and short positions.

- **Financial instruments owned:** As shown in Figure 1, the $312 billion of risky assets mostly includes Treasury and agency bonds ($48 billion), foreign government and agency bonds ($37 billion), mortgage and other asset-backed loans and securities ($11 + $6.5 billion), bank loans ($15 billion), corporate debt securities ($21 billion), equity and convertible debentures ($96 billion), and derivatives ($63 billion). So effectively, the risk factors influencing this portion of the balance sheet include interest rate, currency, equity, and commodities.
- **Balance sheet allocation to business segments:** In terms of balance sheet allocation, most of the $312 billion in financial instruments is attributable to two business segments of Goldman. The segment of Institutional Client Services, which "maintain inventory positions to facilitate market-making in fixed income, equity, currency and commodity products," holds majority ($230 billion) of the financial instruments. The segment of Investing & Lending, whose activities include "investing directly in publicly and privately traded securities and in loans, and also through certain investment funds managed by Goldman," holds $47 billion.[1]

[1] Page 69 of Goldman's 2014 10K.

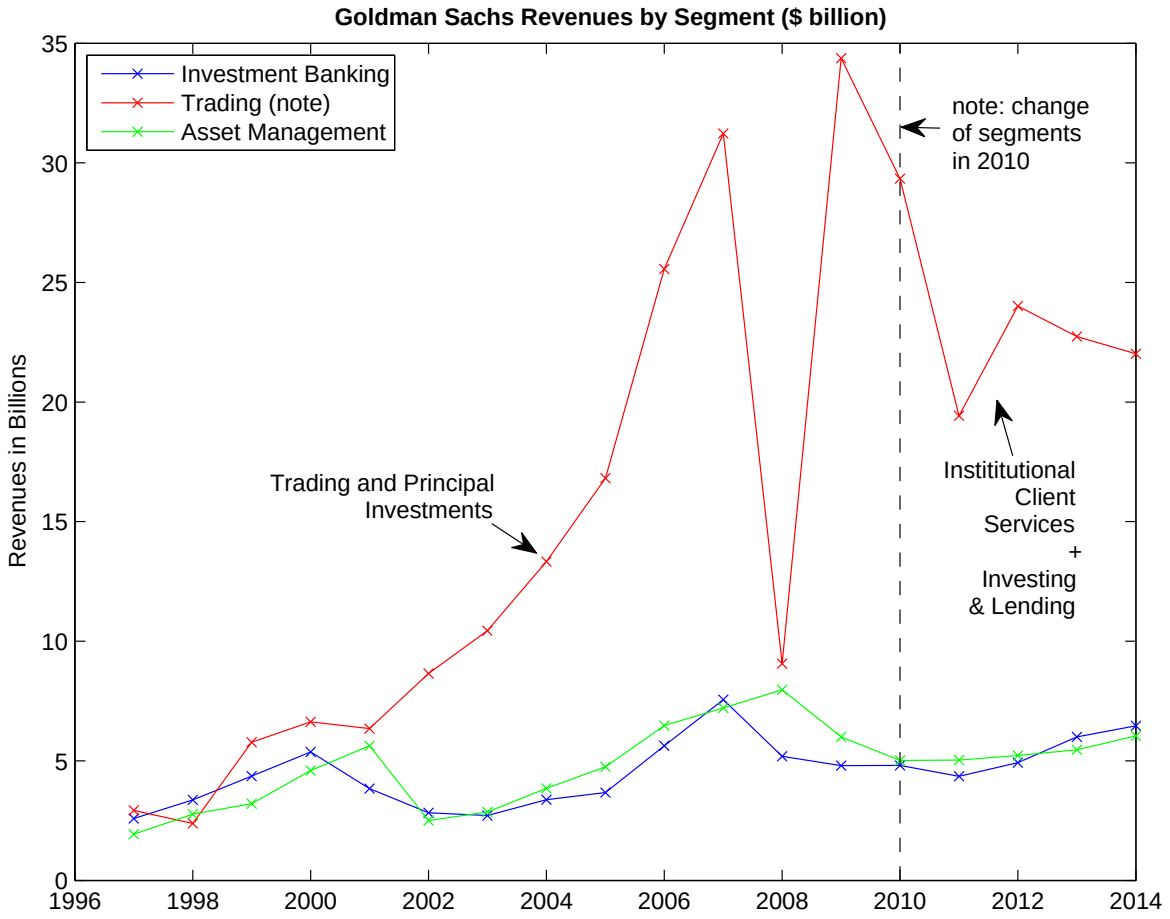


Figure 2: Goldman's Annual Revenues by Business Segment.

From 2009 to 2010, there was a change in how Goldman divide its business segments. My guess is that these two segments belong to the old segment of Trading and Principal Investments. Figure 2 reports the annual revenues by business segment. As you can see, the segment of Trading and Principal Investments (Institutional Client Services + Investing & Lending for post 2010) accounts a large portion of Goldman's revenue and is also the most volatile. Later as we move on to risk management, this segment would be our main focus.

– **Financial instruments sold short:** Figure 1 also reports short positions on financial instruments, valued at $132 billion. On the financial statement, this item shows up in liabilities. For our understanding of the firm's market risk exposure, this item is as important as the $312 billion long positions on financial instruments. It includes short positions on US Treasury and agency bonds ($12 billion), foreign

government and agency bonds ($20 billion), corporate debt securities ($5 billion), equity and convertible debentures ($28 billion) and derivatives ($63 billion). The 10K report does not report the correlation between the risk exposure of the long and the short positions. If the long/short positions are paired as hedging positions, then the net risk exposure will be small. To the extreme, we can say that the net exposure is $312 billion minus $132 billion. Otherwise, we need to take a portfolio approach and take into account of the correlations. More on this later.

– **Derivatives Assets and Liabilities:** The value of derivatives assets is $63 billion and derivatives liabilities is $63 billion, which are sizable positions in relation to Goldman's overall positions in financial instruments. Given the inherent leverage of derivatives, the actual risk exposure per dollar position in these derivatives positions is much higher than the other linear instruments on the list. Again, without knowing the underlying correlations between the derivatives assets/liabilities, it is difficult for us to assess the net exposure. If these derivatives positions are the result of market making activities, then most of the $63 billion assets and liabilities in derivatives will net out and the net exposure will be small.

  Figure 3 gives a more detailed description of Goldman's derivatives positions by major product type on a gross basis. For example, the gross value of interest-rate derivatives totals to $786,362 million in assets and $739,607 million in liability with a total notional amount of $47,112,518 million. As of December 2014, the total notional amount of interest-rate OTC derivatives was $505 trillion, making Goldman an important participant in this market. Compared with the $63 billion derivatives assets and $63 billion derivatives liabilities, these gross value numbers are much larger because they exclude the effects of both counterparty netting and collateral, and therefore are not representative of the firm's counterparty exposure.

  Because of these derivatives positions, Goldman are connected to it many counterparties: financial troubles of its counterparties could have a material impact on Goldman (e.g., AIG in 2008) and Goldman's own financial troubles could have a material impact on its counterparties (e.g., Lehman's default on Lehman's derivatives counterparties). For regulators worrying about financial institutions that are too connected to fail, understanding these derivatives positions should be high on their priority list. After all, the super-senior tranches were a huge cause of concerns during the 2007-08 financial crisis.

– **Liabilities:** According to Table 1, the total liabilities of Goldman in 2014 were

| $ in millions | As of December 2014 Derivative Assets | Derivative Liabilities | Notional Amount |
|---|---|---|---|
| **Derivatives not accounted for as hedges** | | | |
| Interest rates | $ 786,362 | $739,607 | $47,112,518 |
| Exchange-traded | 228 | 238 | 3,151,865 |
| OTC-cleared | 351,801 | 330,298 | 30,408,636 |
| Bilateral OTC | 434,333 | 409,071 | 13,552,017 |
| Credit | 54,848 | 50,154 | 2,500,958 |
| OTC-cleared | 5,812 | 5,663 | 378,099 |
| Bilateral OTC | 49,036 | 44,491 | 2,122,859 |
| Currencies | 109,916 | 108,607 | 5,566,203 |
| Exchange-traded | 69 | 69 | 17,214 |
| OTC-cleared | 100 | 96 | 13,304 |
| Bilateral OTC | 109,747 | 108,442 | 5,535,685 |
| Commodities | 28,990 | 28,546 | 669,479 |
| Exchange-traded | 7,683 | 7,166 | 321,378 |
| OTC-cleared | 313 | 315 | 3,036 |
| Bilateral OTC | 20,994 | 21,065 | 345,065 |
| Equities | 58,931 | 58,649 | 1,525,495 |
| Exchange-traded | 9,592 | 9,636 | 541,711 |
| Bilateral OTC | 49,339 | 49,013 | 983,784 |
| Subtotal | 1,039,047 | 985,563 | 57,374,653 |
| **Derivatives accounted for as hedges** | | | |
| Interest rates | 14,272 | 262 | 126,498 |
| OTC-cleared | 2,713 | 228 | 31,109 |
| Bilateral OTC | 11,559 | 34 | 95,389 |
| Currencies | 125 | 16 | 9,636 |
| OTC-cleared | 12 | 3 | 1,205 |
| Bilateral OTC | 113 | 13 | 8,431 |
| Commodities | — | — | — |
| Exchange-traded | — | — | — |
| Bilateral OTC | — | — | — |
| Subtotal | 14,397 | 278 | 136,134 |
| **Gross fair value/notional amount of derivatives** | $1,053,444 [1] | $985,841 [1] | $57,510,787 |
| **Amounts that have been offset in the consolidated statements of financial condition** | | | |
| Counterparty netting | (886,670) | (886,670) | |
| Exchange-traded | (15,039) | (15,039) | |
| OTC-cleared | (335,792) | (335,792) | |
| Bilateral OTC | (535,839) | (535,839) | |
| Cash collateral netting | (103,504) | (36,155) | |
| OTC-cleared | (24,801) | (738) | |
| Bilateral OTC | (78,703) | (35,417) | |
| **Fair value included in financial instruments owned/ financial instruments sold, but not yet purchased** | $ 63,270 | $ 63,016 | |

Figure 3: Goldman's Derivatives Positions.

at $773 billion, with $167 billion financed by long-term borrowings. The other sources of funding, including unsecured short-term borrowings and Repo financing are mostly short term in nature. Going back to the example of a simple bank, these short-term financings correspond to the demand deposits. Unlike the case of demand deposits, which are FDIC insured, there is no insurance on such short-term funding sources. So some of these short-term financings could evaporate in a moment's notice. Some of the short-term fundings are collateralized (e.g., Repo financing), while some are unsecured (e.g., inter-banking lending or commercial paper).

Table 2: Assets-to-Equity and Financing

| | 2014 | 2010 | 2008 | 2007 |
|---|---|---|---|---|
| assets ($m) | 856,240 | 911,332 | 884,547 | 1,119,796 |
| equity ($m) | 82,797 | 77,356 | 64,369 | 42,800 |
| **assets-to-equity ratio** | 10.3x | 11.8x | 13.7x | 26.2x |
| total liabilities ($m) | 773,443 | 833,976 | 820,178 | 1,076,996 |
| long-term borrowings ($m) | 167,571 | 174,399 | 168,220 | 164,174 |
| other long-term financings ($m) | 7,249 | 13,848 | 17,460 | 33,300 |
| **% of long-term liabilities** | 22.60% | 22.57% | 22.64% | 18.34% |
| total liabilities ($m) | 773,443 | 833,976 | 820,178 | 1,076,996 |
| Repo financing ($m) | 88,215 | 162,345 | 62,883 | 159,178 |
| **% of Repo financing** | 11.41% | 19.47% | 7.66% | 14.78% |

Table 2 shows that in 2014, long-term liabilities account for 22.60% of Goldman's total liabilities, while in 2007, the number was only 18.34%. In recent years, financial firms such as Goldman have experienced disruptions in the credit markets, including reduced access to credit and higher costs of obtaining credit. As such, it is important for them to maintain stable funding in the form of long-term debt. On the other hand, because of the positive term spread (long term yields minus short term yields), long-term financing is more costly.

As we see in Table 2, Repo financing accounted for 11.41% of Goldman's liability in 2014 and 14.78% in 2007. This form of short-term (usually overnight) and collateralized (e.g., Treasury and agency bonds, corporate bonds, and equity) financing is an important source of funding for most investment banks.

– **Leverage:** With total assets at $856 billion, total liabilities at $773 billion, and shareholders' equity at $82 billion, the leverage of a financial firm such as Goldman is markedly different from that of a non-financial firm. As shown in Table 2, the

assets-to-equity ratio was around 10 to 1 in 2014 and 26 to 1 in 2007.

- **Runs on Financial Institutions:** We talked about how a bank run could happen because of the liquidity mismatch between assets and liabilities. After going through the balance sheet of Goldman, it is obvious that the same kind of liquidity mismatch exists in a financial intermediary like Goldman. In particular, long-term liabilities as a percentage of Goldman's total liabilities is 21.67% in 2014 and 15.24% in 2007. In other words, Goldman relies on short-term financing which could evaporate quickly if the markets are no longer confident of Goldman's solvency. Such was the case for Lehman in 2008. After Lehman's default, the solvency of Morgan Stanley and Goldman was seriously questioned by market participants. They had to go out and raise new capital: Morgan Stanley from Japan's Mitsubishi bank on a weekend in the form of a check of $9 billion and Goldman Sachs from Warren Buffett.

  Whenever there is liquidity mismatch in assets and liabilities, there is potential of a run. The 2008 run on money market funds is one such example. Money market funds are an important component of the shadow banking system and are an important source of short-term financing for financial institutions such as Goldman. Money market funds hold commercial paper issued by financial firms such as Goldman and Lehman, and also lend to these dealers in the Triparty Repo market.

  Usually, the assets held by market funds are short term, highly liquid, and of minimum credit risk. This includes short-term Treasury securities and highly rated commercial paper. They mimic bank accounts by allowing check-writing and by fixing the price of a share at $1 – meaning investors could reasonably expect to suffer no losses. Many individual investors keep some cash in money funds, usually in connection with a broader brokerage account. Institutions, including corporations, municipal governments, and pension funds, also find money funds to be a convenient place to park their cash.

  In 2008, one of the money market funds, the Reserve Primary Fund took more risks than many, in an attempt to achieve higher returns and attract more investors. It had invested about $785 million in Lehman's commercial paper, which became worthless after the Lehman default on Monday, September 15, 2008. A run on the fund quickly began, with about $40 billion withdraw (2/3 of the fund's value) by the end of the day on Tuesday.

  The run was not only on this fund alone, it was quickly developing into a run on the entire industry of prime money market funds. In the three weeks between September 10 and October 1, $439 billion would run from the prime funds, while $362 billion

would flow in to the government-only funds (funds invested at least 99.5% in cash, short-term Treasury securities, and Repos collateralized by Treasury securities). This run on money market funds also dried up the commercial paper's market, cutting an important source of short-term funding for financial and non-financial companies.

In 2007 and 2008, we also witnessed the runs on financial institutions such as Bear Stearns, Lehman, Merrill, Morgan Stanley, and even Goldman Sachs. Again, one common characteristics of these firms is the liquidity mismatch between their assets and liabilities. Such firms usually rely heavily on short-term liabilities such as inter-bank lending (Fed Funds and Euro-Dollar), Repo financing (Triparty Repo via money market funds), and commercial paper. Unlike commercial banks such as J.P. Morgan, these firms do not have a broad deposit base. The short-term funding sources they rely upon are subject to runs, especially during financial crises, and the runs on money market funds certainly did not help. Moreover, if a bank was suspected to be the next Lehman, it would have even more trouble funding itself through the short-term funding sources in Fed Funds, Repo, or commercial paper. At the same time, its long-term assets are deteriorating and its counterparties are requesting for more collateral for existing liabilities connected with derivatives positions.

As you've read in the popular press, it has been a death spiral in real time. By the way, the Mitsubishi story was in Andrew Ross Sorkin's book on "Too big to fail," which reads like a thriller (if you are looking for entertainment on a weekend).

# 2 Market Risk Measurement

- **Value-at Risk:** For financial institutions, the larger economic consequences of market risk are felt over relatively short time horizons, often over a few weeks, if not days. Discussions between regulators and their constituent financial institutions have results in a widely applied measure of market risk called value-at-risk.

  For a portfolio of securities (long and short positions), VaR is the potential loss in value due to adverse market movements over a defined time horizon with a specified confidence level.

  - **The scope of the VaR calculation:** Going back to the Goldman's balance sheet, the items listed in Figure 1 will be the scope over which the VaR calculation is done. Moreover, only those financial instruments in Goldman's trading book will be included in the VaR calculation while financial instruments held in

Goldman's banking book are excluded from the VaR calculation. The firm has the discretion in choosing where to allocate a security: to its trading book or banking book. Securities in the banking book are held to maturity, while those in trading books are more frequently traded. So the VaR calculation will cover only the portion of the financial assets listed in Figure 1 that are allocated to the bank's trading book.

- **Confidence level and time horizon:** The typical confidence level $p$ is 99% or 95%, focusing on the 1% or 5% worst-case scenario. To go further out in the tail, sometimes banks also calculation VaR with a confidence level of 99.6%, which is linked to the 0.4% worst-case scenario.

  For a typical broker-dealer or proprietary trading operation, the larger economic consequences of market risk are felt over relatively short time horizons. So the typical time horizon is over two weeks (10 days) or one day.

- **Goldman's VaR:** For both risk management purposes and regulatory capital calculations, Goldman uses a single VaR model which captures risks including those related to interest rates, equity prices, currency rates and commodity prices. The VaR used for regulatory capital requirements (regulatory VaR) differs from risk management VaR due to different time horizons and confidence levels: 10-day and 99% for regulatory VaR and one-day and 95% for risk management VaR. These two VaR calculations also differ in the scope of positions on which VaR is calculated. For our analysis, we will focus on the VaR reported by Goldman for risk management purpose: one-day and 95%.

- **Calculating VaR:** The original intention of the VaR measure is to capture the tail events: the amount of portfolio loss when a 5% left-tail event happens over a day. The actual implementation of the VaR measure, however, relies heavily on the assumption of a normal distribution.

  Let's start with a simple example of a portfolio consisting entirely of the S&P 500 index. Suppose that the current market value of the portfolio is \$100 million. Using the historical return data available up to day $t$, the EWMA model gives us a volatility forecast $\sigma_{t+1}$ for the next day's stock return $R_{t+1}$. Standing at day $t$, the value of the portfolio at the end of day $t+1$ would be $\$100\,\mathrm{M} \times (1 + R_{t+1})$. As discussed in the volatility class, the mean of $R_{t+1}$ is negligible for this one-day horizon. So let's focus on the impact of volatility on the profit/loss of this portfolio.

  Focusing on the potential loss, we are interested in how much we would lose if a 5%

tail event happens. Assuming normal distribution, a 5% tail corresponds to a critical value of $-1.645\sigma$; a 1% tail corresponds to a critical value of $-2.326\sigma$. Using these number, the loss in portfolio value associated with a 5% worst-case scenario would be

$$\text{VaR} = \$100\,\text{M} \times 1.645 \times \sigma_{t+1}$$

For daily returns on the S&P 500 index, the volatility is about 1%. So VaR= \$1.645 M.

As you can see, although VaR was designed to capture the tail events, the actual implementation of VaR uses a normal distribution. As a result, calculating VaR bolts down to calculating volatility:

$$\text{VaR} = \text{portfolio value} \times 1.645 \times \text{daily portfolio sigma}\,.$$

Moreover, given how VaR is phrased, one might mistaken VaR as a predictor of the future. In practice, VaR is a measure of the past because the portfolio volatility is estimated using historical returns. In fact, if you calculate the VaR for a risky portfolio right before the any of the crisis, you will not be able to pick up anything above and beyond what the volatility estimate can give you. In this sense, VaR is a more reactive measure: reacting to market volatility.

Figure 4 plots the time-series of VaR for a hypothetical portfolio consisting entirely of the S&P 500 index. Suppose that this portfolio has a market value of \$100 million on January 2, 2008. For comparison, I also plot the time-series of the daily EWMA volatility of the S&P 500 index multiplied by 1.645. On January 2, 2008, two time-series started at the same level because VaR = $\$100M \times 1.645 \times$ Sigma.

As shown in Figure 4, the time variation of VaR has two driving forces: the market value of the portfolio and the portfolio volatility. As the year progressed, this passively managed portfolio kept losing its market value. As a result, the blue line (VaR) is lower than the red line (volatility). By late October and early November, it is obvious that the portfolio has lost quite a bit of its value because the difference between the blue line and the red line became quite large. Overall, however, it is obvious that the time variation in VaR tracks the volatility movement quite closely.

- **Calculating VaR for a Portfolio:** As shown in Figure 1, the trading portfolio of a large financial intermediary such as Goldman could be large and complex. In one of its 10K forms, Goldman mentioned 6 million individual positions, 70,000 market factors and 1 million computing hours in its risk management calculations:

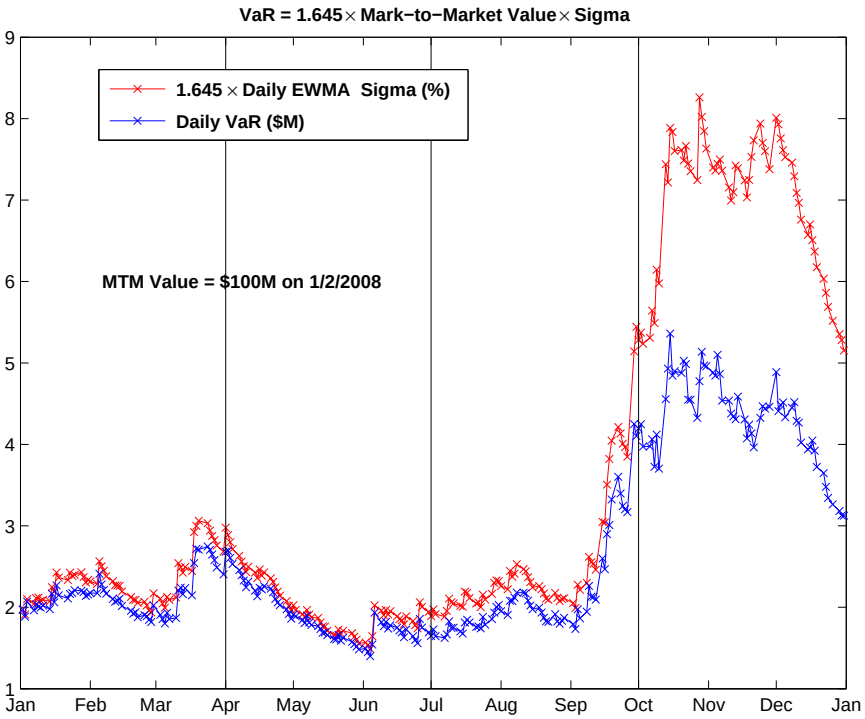


Figure 4: Time series of daily VaR for a portfolio of the S&P 500 Index with an initial market value of $100 million on Jan. 2, 2008.

*"We also rely on technology to manage risk effectively. While judgment remains paramount, the speed, comprehensiveness and accuracy of information can materially enhance or hinder effective risk decision making. We mark to market approximately 6 million positions every day. And, we rely on our systems to run stress scenarios across multiple products and regions. In a single day, our systems use roughly 1 million computing hours for risk management calculations.*

*When calculating VaR, we use historical simulations with full valuation of approximately 70,000 market factors. VaR is calculated at a position level based on simultaneously shocking the relevant market risk factors for that position. We sample from 5 years of historical data to generate the scenarios for our VaR calculation. The historical data is weighted so that the relative importance of the data reduces over time. This gives greater importance to more recent observations and reflects current asset volatilities, which improves the accuracy of our estimates of potential loss. As a result, even if our inventory positions were unchanged, our VaR would increase with increasing market volatility and vice versa."*

- **Risk Factors:** The first task of a risk manager is to identify risk factors that are important for risk management purposes. Suppose there are $N$ risk factors. For this $N$ risk factors, the risk manager calculates the covariance-covariance matrix using the EWMA approach. On day $t$, $\Sigma_{t+1}$ is the covariance-variance matrix calculated using return data up to day $t$:

$$\Sigma_{t+1} = \begin{pmatrix} (\sigma_1)^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \dots & \rho_{1N}\sigma_1\sigma_N \\ \rho_{21}\sigma_2\sigma_1 & (\sigma_2)^2 & \rho_{23}\sigma_2\sigma_3 & \dots & \rho_{2N}\sigma_2\sigma_N \\ \rho_{31}\sigma_3\sigma_1 & \rho_{32}\sigma_3\sigma_2 & (\sigma_3)^2 & \dots & \rho_{3N}\sigma_3\sigma_N \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{N1}\sigma_N\sigma_1 & \rho_{N2}\sigma_N\sigma_2 & \rho_{N3}\sigma_N\sigma_3 & \dots & (\sigma_N)^2 \end{pmatrix},$$

  where $\rho_{ij}$ is the correlation between risk factor $i$ and $j$ and $\sigma_i$ is the volatility for risk factor $i$. To simplify the notation, I dropped the time-subscripts for $\rho$ and $\sigma$, which are EWMA estimates using data up to time $t$ and time-stamped by $t+1$.

- **Risk Mapping:** Given the $N$ risk factors, the next step is to map the individual positions in the firm's portfolio into positions on the risk factor. For example, a \$100 million position in AAPL maps to \$100 million position in the risk factor for the US equity market. After this risk mapping is done, the risk manager will

have a vector of portfolio weights on day $t$:

$$W_t = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \cdots \\ w_N \end{pmatrix},$$

where $w_i$ is the portfolio weight associated with risk factor $i$. Again, I dropped the time subscripts for $w_i$ to simplify the notation.

– **Portfolio Volatility and VaR:** Armed with the variance-covariance matrix $\Sigma$ and the portfolio weights $W$, the portfolio volatility can be calculated using the matrix operation:

$$\sigma^2_{t+1} = W'_t \times \Sigma_{t+1} \times W_t\,,$$

where $W'_t$ is the transpose of $W_t$. This might be a good time for you to get yourself familiar with matrix operations such as mmult and transpose in Excel. Once the portfolio volatility is obtained, the portfolio VaR is

$$\text{VaR} = \text{portfolio value} \times 1.645 \times \text{daily portfolio sigma}\,.$$

If we are interested in calculating VaR for positions related only to interest rates, we can construct an interest rate portfolio weight $W^{\text{IR}}$ by turning off the portfolio weights on other risk factors (i.e., making the weights zero). We can then calculate the volatility associated with only the interest rate exposure:

$$\left(\sigma^{\text{IR}}_{t+1}\right)^2 = \left(W^{\text{IR}}_t\right)' \times \Sigma_{t+1} \times W^{\text{IR}}_t$$

- **Goldman's VaR, Magnitude:** Table 3 reports Goldman's daily average VaR for a few selected years. Goldman also reports VaR separately for the risk exposures in interest rates, equity, currency, and commodities. As you can see, the individual VaR's do not add up to equal to the total VaR because of the diversification effect. Only when these four risk factors are perfectly correlated, would we expect to see the four individual VaR's to sum up to equal to the total VaR.

  The VaR numbers for Goldman are in the range of $100 million. Recall that in calculating the these VaRs, the key ingredients are the portfolio value and the portfolio

Table 3: Goldman's Average Daily VaR

| **Financial Instruments** | | | | |
|---|---|---|---|---|
| in millions | 2014 | 2010 | 2008 | 2007 |
| Long | 312,248 | 356,953 | 328,325 | 452,595 |
| Short | 132,083 | 140,717 | 175,972 | 215,023 |
| Long - Short ($m) | 180,165 | 216,236 | 152,353 | 237,572 |
| **Average Daily VaR** | | | | |
| in millions | 2014 | 2010 | 2008 | 2007 |
| Total | 72 | 134 | 180 | 138 |
| Interest Rates | 51 | 93 | 142 | 85 |
| Equity Prices | 26 | 68 | 72 | 100 |
| Currency Rates | 19 | 32 | 30 | 23 |
| Commodity Prices | 21 | 33 | 44 | 26 |

volatility:

$$\text{VaR} = \text{portfolio value} \times 1.645 \times \text{daily portfolio sigma}\,.$$

If we know one of them, then knowing VaR can help us back out the other. The problem is that neither the portfolio value or the portfolio sigma is reported by Goldman. Still, let's do some guess work.

Let's first suppose that the long/short positions are paired positions and the net exposure is long minus short. So for 2014, the number is $180,165 million. Suppose that 10% of these positions have been allocated by Goldman to its trading book and fall under the scope of VaR calculation. So portfolio value = $18 billion. Then

$$\text{daily portfolio sigma} = \frac{\text{VaR}}{\text{portfolio value} \times 1.645} = \frac{\$72}{\$18,016.5 \times 1.645} = 24\,\text{basis points}\,.$$

Repeat the same exercise for 2007 (again assuming the trading portfolio consists only 10% of the long-short positions):

$$\text{sigma} = \frac{\text{VaR}}{\text{portfolio value} \times 1.645} = \frac{\$138}{\$23,757.2 \times 1.645} = 35\,\text{basis points}\,.$$

For 2008, the inferred volatility is higher, around 72 basis points. For 2010, it is 31 basis points.

To assess these levels of daily volatility, let's compare them with numbers that we are

familiar with. As you know, the equity market has a daily volatility around 100 basis points. For the fixed income market, the standard deviation of the daily changes in the 10-year yields is around 7 basis points. Assuming a duration of 8 years for 10-year bonds, the daily volatility of 10-year treasure bond is about 56 basis points. The typical annual volatility for a currency portfolio is about 9%, making the daily volatility of a currency portfolio at about 57 basis points.

Now back to our inferred volatility of around 24 basis points in 2014, which seems low compared to the numbers we are familiar with. There could be several reasons for this. The diversification benefit across the asset classes will further reduce the overall portfolio volatility. The hedging activities within the trading book will reduce the portfolio volatility. Finally, it is also possible that the trading book is smaller than the 10% assumption we made earlier. Or it could be that the trading book of Goldman is of very low volatility. In any case, this is not meant to be a serious exercise looking into the trading book of Goldman.

- **More on the Portfolio:** In estimating the portfolio value of Goldman, we assumed that the long/short positions are paired and think of the net exposure as long minus short. Let's do a little better than that.

  Using the 2014 number, it is long \$312B and short \$132B. So the portfolio weight on the long portfolio $R_t^L$ is $w^L = 312/(312 - 132) = 173\%$, the weight on the short portfolio $R_t^S$ is $w^S = -73\%$, and the total portfolio is

$$R_t = w^L\, R_t^L + w^S\, R_t^S = 173\%\, R_t^L - 73\%\, R_t^S$$

  The volatility of the portfolio is

$$\mathrm{var}(R_t) = (w^L)^2\, \mathrm{var}(R_t^L) + (w^S)^2\, \mathrm{var}(R_t^S) + 2\rho\, w^L\, w^S\, \mathrm{std}(R_t^L)\, \mathrm{std}(R_t^S)\,,$$

  where $\rho$ is the correlation between these the long and short portfolios. It is difficult for us to assess the magnitude of $\rho$ without seeing the book. So let's think of different scenarios.

  Suppose $\rho = 1$ and $\mathrm{std}(R_t^L) = \mathrm{std}(R_t^S) = \sigma$, the volatility of the portfolio becomes

$$\mathrm{var}(R_t) = (w^L)^2\, \sigma^2 + (w^S)^2\, \sigma^2 + 2 w^L\, w^S\, \sigma^2 = (w^L + w^s)^2 \sigma^2\,.$$

  We are back to the earlier assumption that the long/short positions are paired and the

net exposure is $312 billion minus $132 billion.

Suppose $\rho$ is not 1 but close to one. It is very likely that there are hedging activities between the long/short portfolios, but the hedging will not take out all of the risk. As a result, the portfolio volatility would be higher because of the leverage involved in the long/short portfolio.

Take the extreme case of $\rho = 0$, and again assuming $\text{std}(R_t^L) = \text{std}(R_t^S) = \sigma$, the portfolio volatility is $\sqrt{(w^L)^2 + (w^S)^2}\sigma$, which is $1.88\sigma$ for the 2014 case. This is not surprising because leverage increases portfolio volatility. Again, these are not meant to be a serious investigation into the trading book of Goldman. Rather, I am using them as useful exercises on calculating the volatility of a portfolio.

- **Goldman's VaR, Time-Variation:** Let's also take a look at the time-variation of Goldman's VaR to see if there is anything we can learn. Figure 5 plots Goldman's daily VaR in 2008 (from Goldman's 10K), along with the VaR of a hypothetical portfolio consisting entirely of the S&P 500 index. I set the hypothetical portfolio to have an initial market value of $8 billion so that the portfolio VaR at the beginning of the year matches the VaR number for Goldman's portfolio. Of course, unlike the passive portfolio in the S&P 500 index, the Goldman's portfolio is actively managed and most likely, the positions were adjusted to the market conditions at the time.

  As shown in Figure 5, for 2008, the Goldman's VaR bottomed to a level close to $130 million in mid-February (with a visible spike in mid-January). In the last quarter of 2008, Goldman's VaR peaked to a level around $240 million. As discussed earlier, the VaR of a portfolio increases for two reasons: increasing portfolio volatility or increasing market value of the portfolio. Overall, it is difficult for us to learn too much from this time-series plot of Goldman's VaR. The visible spike in mid-January was interesting (no significant increase in the stock market volatility on the same day), and was probably due to a sudden increase in Goldman's portfolio volatility.

  For a risk manager, sudden spikes in VaR could be alarming as well as informative. It was reported in the media that in December 2006, Goldman's various indicators, including VaR and other risk models, began suggesting that something was wrong. Not hugely wrong, but wrong enough to warrant a closer look. As a result of that effort, Goldman started to reduce their exposure to mortgage-back securities in late 2006.

  In a large financial firm such as Goldman, trading and market-making take place in a decentralized fashion on various trading desks. In calculating the VaR number, individual positions scattered in different parts of the firm are aggregated and compiled

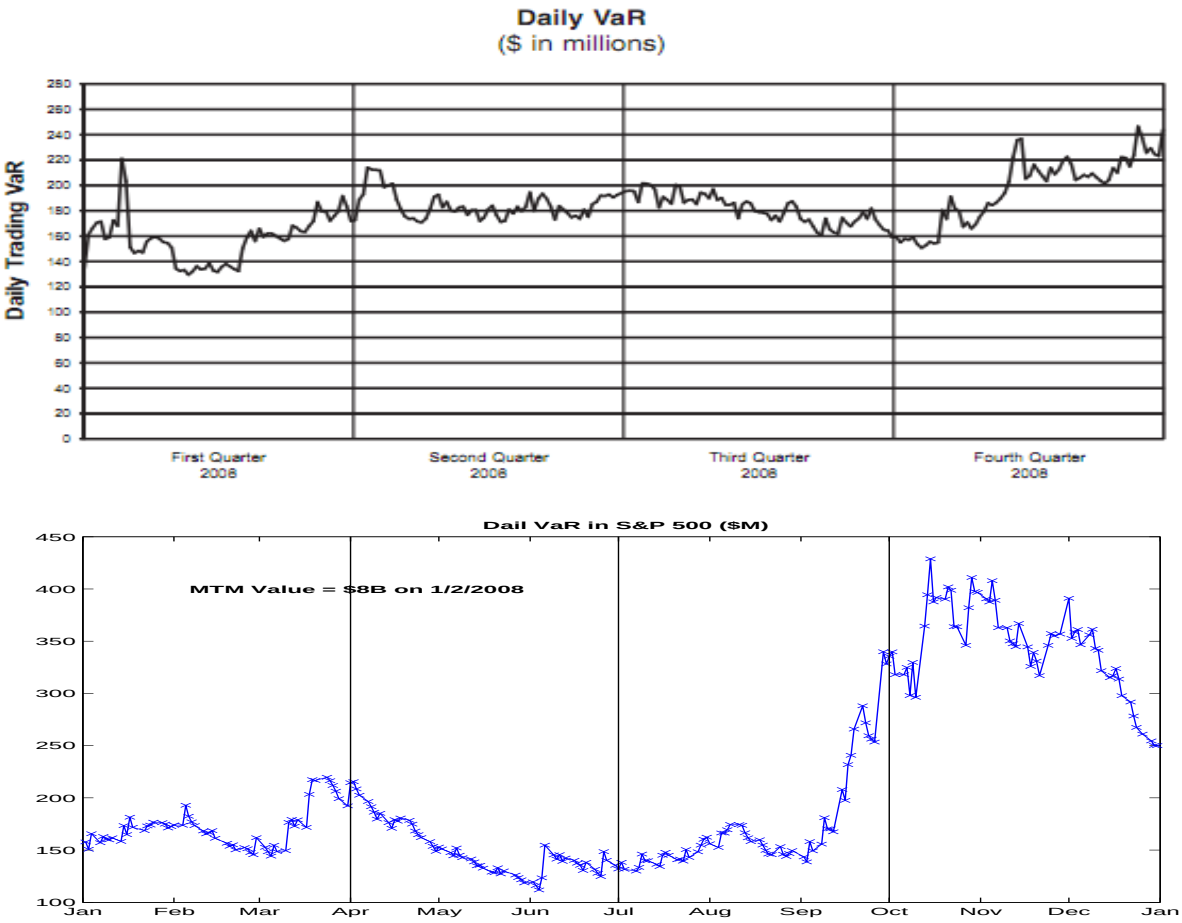


Figure 5: Time-Series of Daily VaR of Goldman Sachs in 2008 vs. Daily VaR of $8 billion in the S&P 500 Index in 2008

into one large portfolio. At the market close, executives of the firm have information of the firm's overall portfolio value as well as its loss and profit from the days before; the portfolio volatility as well as its increase or reduction from the days before. This effort itself is meaningful for the firm, and how to make the VaR measure useful relies crucially on the judgment of a good risk manager.

It would be naive for a risk manager to believe that a VaR of $100 million means that the potential portfolio loss (of a 5% worst-case scenario) is somehow in the neighborhood of $100 million. If this is how VaR is being used in practice, then, quoting the hedge fund manager David Einhorn, VaR is *"relatively useless as a risk-management tool and potentially catastrophic when its use creates a false sense of security among senior managers and watchdogs. This is like an air bag that works all the time, except when you have a car accident."*

- **Days Exceeding VaR:** On each business day, Goldman compares its daily trading net revenues with the VaR calculated at the end of the prior business day and report, in each year's 10K form, the number of days the firm incurs trading losses in excess of the 95% one-day VaR. Figure 6 plots this VaR exception from 1999 through 2014. As a comparison, the VaR exception numbers for a hypothetical portfolio of the S&P 500 index are also plotted in Figure 6.

  Let's start with bottom panel of Figure 6. Given the definition of 95% VaR, the expectation is that the VaR limits would be exceed 5% of the days in a year: $5\% \times 252 = 12.6$. In some years, because of the tail fatness, the days of VaR exception were above 12.6 days (e.g., 2007 and 2008). In general, the numbers fluctuate around 12.6 days per year. The top panel reports the days of VaR exception for Goldman. The results are quite peculiar: most of the years, the numbers were either 0 or 1. Only during the 2007-08 crisis, did these numbers became meaningfully large.

# 3 Regulatory Requirements

The regulatory requirements for banks makes a very long list and requires exhaustive and patient learning. The landscape of regulatory requirements is still in transition with new rules and requirements phasing in over the next few years. Their effectiveness remains to be evaluated. In the meanwhile, the increasing regulatory requirements have certainly created more risk compliance jobs.

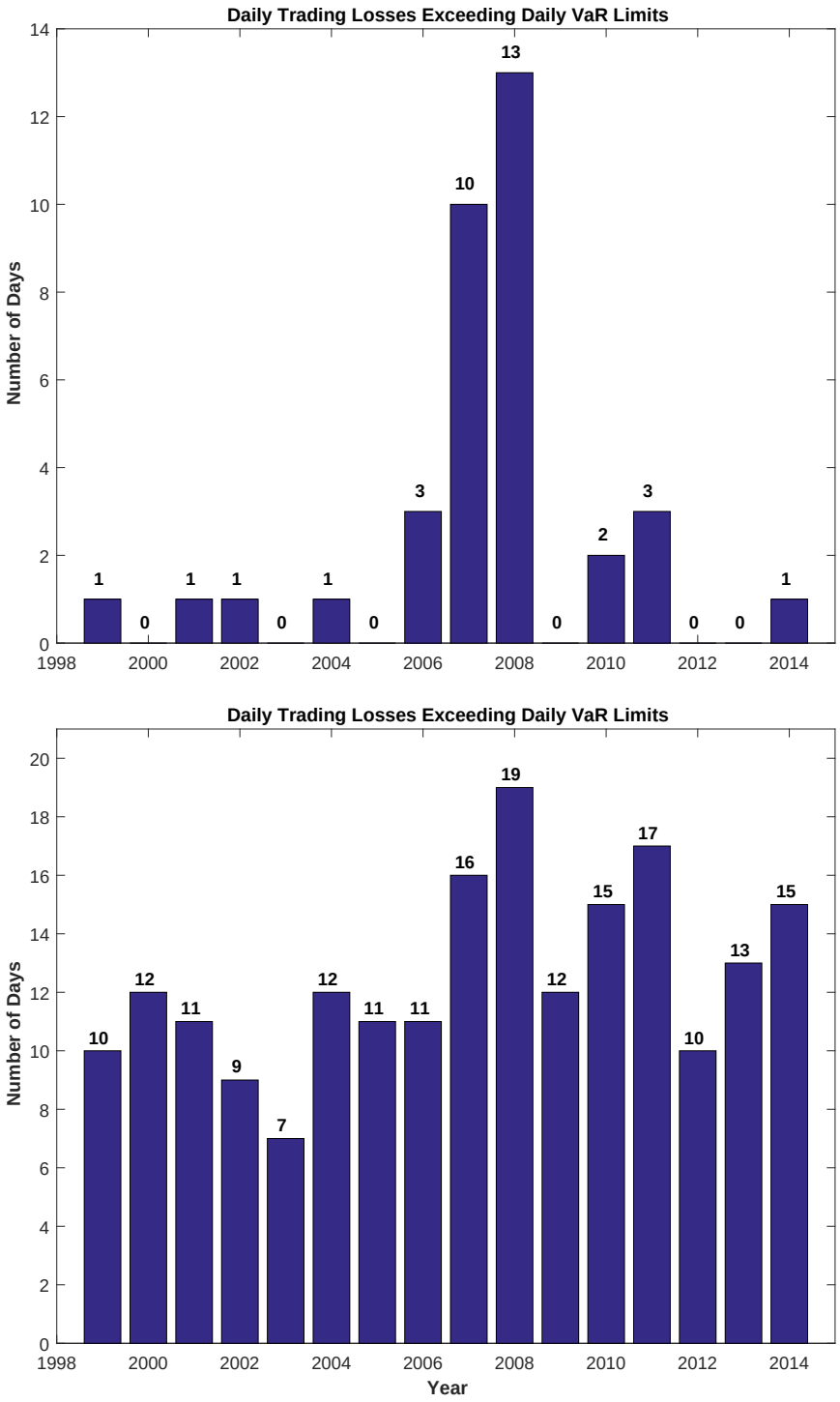


Figure 6: The Number of VaR Exception Days per Year.

- **Capital Adequacy:** As we learned in the example of a simple bank, equity acts as a buffer to cushion the downfall a bank during stressful situations. An important component of the regulatory requirements is expressed as capital ratios that compare measures of regulatory capital to risk weighted assets (RWAs). Capital ratios are ratios of Capital to Assets. Let's take a look at the regulatory measures of these two items separately.

- **Risk Weighted Assets:** Going back to our simple example, the bank holds 10 dollars in cash and 90 dollars in risky loans. For regulatory purpose, the 10 dollars in cash is safe and carries a zero weight in RWA. For the risky loans, there are two kinds of risks: credit and market risk. The bank incurs credit risk because the firms the bank lends to might default. The associated risk weights depend on the type of counterparty (e.g., sovereign, bank, broker-dealer or other entity), the credit worthiness of the counterparty (Aaa, A, Baa, etc), and whether or not the loan is collateralized. In the case of the loan, the bank also incurs market risk because the fluctuations of interest rates. If the bank also holds equity or loans in foreign currencies, then stock market risk and currency risk will also affect the bank's asset. Overall, the bank's RWAs is the sum of its credit RWAs and market RWAs, and most of the regulatory capital ratios are calculated as a ratio to this RWA number.

  Figure 7 reports Goldman's RWA in 2014, which including three components, credit, market, and operational RWAs. The actual calculations of these number requires some training, which I am not at all an expert. But the Table is a good starting point for us to understand the various components of RWA and their relative importance.

  From Figure 7, we can see that the regulatory landscape is still in transition. For example, Goldman reported its RWAs in 2014 under two sets of capital frameworks: Basel III Advanced Rules and Standardized Capital Rules.

- **Regulatory Capital and Capital Ratios:** There are also various ways of measuring regulatory capital, including the newly proposed Common Equity Tier 1 (CET1) capital. Figure 8 is a good starting point to understand the differences in these capital measures. Essentially, what matters in capital requirement is the quantity as well as quality of the capital.

  Capital requirements are expressed as capital ratios of the various regulatory capitals to RWAs. Figure 9 the minimum ratios under the Revised Capital Framework as of December 2014 and January 2015, as well as the minimum ratios that expected by Goldman to apply at the end of the transitional provisions beginning January 2019.

| $ in millions | As of December 2014 Basel III Advanced | Standardized |
|---|---|---|
| **Credit RWAs** | | |
| Derivatives | **$122,501** | **$180,771** |
| Commitments, guarantees and loans | **95,209** | **89,783** |
| Securities financing transactions [1] | **15,618** | **92,116** |
| Equity investments | **40,146** | **38,526** |
| Other [2] | **54,470** | **71,499** |
| **Total Credit RWAs** | **327,944** | **472,695** |
| **Market RWAs** | | |
| Regulatory VaR | **10,238** | **10,238** |
| Stressed VaR | **29,625** | **29,625** |
| Incremental risk | **16,950** | **16,950** |
| Comprehensive risk | **8,150** | **9,855** |
| Specific risk | **79,918** | **79,853** |
| **Total Market RWAs** | **144,881** | **146,521** |
| **Total Operational RWAs** | **97,488** | **—** |
| **Total RWAs** | **$570,313** | **$619,216** |

1. Represents resale and repurchase agreements and securities borrowed and loaned transactions.

2. Includes receivables, other assets, and cash and cash equivalents.

Figure 7: Credit, Market, and Operational Risk Weighted Assets Reported by Goldman Sachs.

| *$ in millions* | As of December 2014 |
|---|---|
| Common shareholders' equity | $ 73,597 |
| Deductions for goodwill and identifiable intangible assets, net of deferred tax liabilities | (2,787) |
| Deductions for investments in nonconsolidated financial institutions | (953) |
| Other adjustments | (27) |
| **Common Equity Tier 1** | 69,830 |
| Perpetual non-cumulative preferred stock | 9,200 |
| Junior subordinated debt issued to trusts | 660 |
| Other adjustments | (1,257) |
| **Tier 1 capital** | 78,433 |
| Qualifying subordinated debt | 11,894 |
| Junior subordinated debt issued to trusts | 660 |
| Other adjustments | (9) |
| **Tier 2 capital** [1] | 12,545 |
| **Total capital** | $ 90,978 |

Figure 8: Regulatory Capital.

| | December 2014 Minimum Ratio [1] | January 2015 Minimum Ratio [1] | January 2019 Minimum Ratio |
|---|---|---|---|
| CET1 ratio | 4.0% | 4.5% | 8.5% [4] |
| Tier 1 capital ratio | 5.5% | 6.0% | 10.0% [4] |
| Total capital ratio | 8.0% [3] | 8.0% [3] | 12.0% [4] |
| Tier 1 leverage ratio [2] | 4.0% | 4.0% | 4.0% |

Figure 9: Minimum Capital Ratios and Capital Buffers.

The framework of RWA has been subject to much criticism, especially given that banks are allowed to use their own risk models to calculate Market RWAs. Tier 1 leverage ratio moves away from the RWA framework, and measures the ratio of Tier 1 capital to the average adjusted total assets.

- **Liquidity Adequacy:** We also learned in our simple example that the liquidity mismatch between the assets and liabilities is the root cause of runs on financial institutions. The more recent regulatory effort in Basel III pays special attention to this liquidity issue and proposed two liquidity measures.

  - **LCR:** The measure of Leverage Coverage Ratio (LCR) is to promote the short-term resilience of the liquidity risk profile of banks. It does so by ensuring that banks have an adequate stock of unencumbered high-quality liquidity assets that can be converted easily and immediately in private markets into cash to meet their liquidity needs.
  - **NSFR:** Another proposed measure in Basel III is Net Stable Funding Ratio (NSFR), which requires that long-term financing resources (e.g., equity and any liability maturing after one year, retail deposits, deposits from non-financial corporates and public entities) must exceed long-term commitments.

- **The Last Taxi Cab in the Train Station:** I heard this story from Prof. Doug Diamond who was the Fischer Black Visiting Professor of Finance at MIT Sloan in 2015.

  On a cold and rainy night, the last train arrived at a small town in ... Europe. There was just one passenger getting off from the train and he is tired and hungry and eager to go home. There was one taxi cab waiting at the train station. The passenger got in and asked to be taken to his home, which is only a few miles away from the train station. But the taxi driver told him that he cannot take him there. According to the local law, there must always be one taxi cab waiting at the train station.

  It is one of those story that sounds crazy and yet not totally crazy. Going back to the regulatory requirements on capital and liquidity adequacy, it is possible that banks are required to hold liquidity that goes unused, just like the last taxi cab at the train station. But this does not necessarily mean that the unused liquidity was not useful. In a way, the presence of the unused liquidity deters the run on the financial institution. For a bank, the calculation would be how costly it is to hold the unused liquidity vs the cost of a run. For regulators, the concern is not on just one bank, but the liquidity

and stability of the entire financial system. As such, they would want to focus on the liquidity adequacy of those highly connected financial institutions.

Jun Pan
junpan@mit.edu

253-3083
E62-624

## Class 19: Fixed Income, Yield and Duration

This Version: November 28, 2016[1]

# 1 From Equity to Fixed Income

- **Vehicles for Risk:** Moving from equity to options to bonds and, later, to OTC derivatives, there is always one thing in common: each market is a vehicle for risk. The nature and origin of the risk might vary from one market to the other, but our approach to risk remains the same.

  We plot the time-series data to see how it varies over time. We map the historical experiences into a distribution and use it as a basis to envision future scenarios. Thinking of the future in a static fashion as one fixed future date, we employ random variables to model the distribution at this future date (e.g., the CAPM). Thinking of the future in a dynamic fashion as a path leading into the future, we use stochastic processes to model the random paths (e.g., Black-Scholes). Either way, we use these models to price the risk involved, taking into account not only the likelihood and magnitude of the risk, but also investors' attitudes to the risk. After this is done, we go back to the data to see how well our model performs. Very often, the data surprises us. In this process of model meeting the data, new insights arise.

- **Relating one to the other:** You might also notice that, in Finance, we keep ourselves busy by relating one thing to the other. For example, in the equity market, we relate the individual stock returns $R_t^i$ to the contemporaneous returns of the market portfolio $R_t^M$. The pricing of an individual stock is done through the pricing of the market:

$$E(R_t^i) - r_f = \beta^i \left( E(R_t^M) - r_f \right) .$$

[1]This note was originally written in November 2015. I have not had the chance to update it for Fall 2016. In many places, "right now" means Fall 2015. Just a quick update on the numbers: as of November 16, 2016, the three-month Treasury yield is at 46 basis points, the 10-year yield is at 2.22%, and the 30-year at 2.92%. On November 8, 2016, the 10-year was at 1.88%, followed by 2.07%, 2.15%, and 2.23% on November 9, 10, and 14.

By doing so, we narrow our attention down to one risk factor: the market portfolio. In the crowd of thousands of stocks, your eyes are on this one and one thing only, and everything else fades into the background.

In options, we relate the time-$t$ option price $C_t$ to two things: the price of the underlying stock $S_t$ and the volatility of the underlying stock $\sigma$. The relation between $C_t$ and $S_t$ is useful, but what really makes options unique is the relation between $C_t$ and $\sigma$. This is especially important when we step outside of the Black-Scholes model and allow $\sigma_t$ to vary over time: now options are unique vehicles for the risk in $\sigma_t$. This is why I asked you to pay special attention to this approximation for an ATM option:

$$C_t/S_t = P_t/S_t \approx \frac{1}{\sqrt{2\pi}}\,\sigma\sqrt{T}\,.$$

Now we are studying the fixed-income market, which is large and important, encompassing products such as Treasury bonds ($12.5tn), mortgage-backed securities ($8.7tn), corporate bonds ($7.8tn), Muni ($3.6tn), money market funds ($2.9tn), agency bonds ($2.0tn), and asset-backed securities ($1.3tn). The numbers in parentheses are amount outstanding as of end 2014. At the center of our attention is the risk that is common to all of these products: interest rate fluctuations. Not one interest rate, but many: one for each maturity. Putting them together, we have a yield curve. In Finance, there is no other risk that is more important than this yield curve risk. It is fundamental to everything we do in Finance. It is the basis from which all other discount rates are calculated.

In dealing with this risk, we prefer to work in the yield space because it is more convenient, but the profit/loss happens in the dollar space. As a result, we will be busy relating one thing to the other again. This gives rise to concepts such as duration and convexity. An outsider might look at these funny names and accuse people in Finance of creating unnecessary concepts so as to confuse and take advantage of those who know less about finance. There might be such practices going on elsewhere on Wall Street, but concepts such as duration and convexity and Black-Scholes implied vol are created out of necessity. I cannot imagine myself navigating the bond market without having tools like duration and convexity.

- **Focus on What's Important:** In talking about beta in equity, implied-vol in options, and duration and convexity in bonds, my intention is to remind you to focus on what's important.

Often, I notice that some students have the tendency to focus on the small and trifling things first before trying to digest the more important message. When you look at a tree, your attention goes first to the overall structure and shape, not to a small offshoot from a branch of the tree (unless there is a cat sitting there). If you are drowning, you grab the nearest and largest lifesaver available; you don't stop to examine the color or the make of the lifesaver. Nor do you question whether or not the lifesaver is made of sustainable materials.

So please, go for the important concept first. Only after you understand these concepts really well, then you have the luxury in digging into the minute details. Of course, ideally, you would like to be good at both: big-picture and rigor. But in the process of learning, it makes sense to go after the big picture first.

While I am on this topic, let me also add that you should always bring your common sense back to anything you do in Finance. For example, it is very easy to get lost when working on a project. Sooner or later, the model and the spreadsheet become the boss and you the slave. Use your common sense. Don't invest in any fancy models or techniques until you have a very clear view of why you need them. Otherwise, it will be garbage in and garbage out. In the process, you might manage to impress yourself and a few others with the fancy techniques and models. But in truth, it is mostly confusion.

The same thing applies to a professor. If, after each class, I make you more confused than before, then I am not doing a good job in teaching the materials. That is why I am writing the lecture notes, to give myself ... a second chance.

- **In the Return Space:** Coming back to our main topic, I list in Table 1 summary statistics of equity (the CRSP value-weighted index) and bond returns using monthly data from 1942 through 2014. In the second panel of the table, I also report the numbers for the more recent period from 1990 through 2014.

  For the sample period from 1942 through 2014, the average monthly return of the US stock market is 1.03% and the volatility is about 4.16%. In annualized terms, the average return is 12.33% and the volatility is 14.4%. (The 20% annual volatility number we've been using includes the great depression.) For the same period, the average return of a 10-year bond is about 47 basis points per month and the volatility is about 2%. Not surprisingly, with decreasing maturity (and duration), both the average return and volatility decrease for shorter maturity bonds. The one-month TBill has an average return of 32 basis points per month, and an average yield of

$0.32\% \times 12 = 3.84\%$. The monthly volatility of the one-month Treasury bill is 0.26%, which is only a small fraction of that in the stock market (4.16%).

Table 1: Monthly Equity Returns and Bond Returns

| **Monthly** | mean | std | Sharpe | min | max | correlation with | | |
|---|---|---|---|---|---|---|---|---|
| **1942-2014** | (%) | (%) | ratio | (%) | (%) | Stock | TBill | 10Y |
| Stock | 1.03 | 4.16 | 0.17 | -21.58 | 16.81 | 1.00 | -0.05 | 0.10 |
| 10Y Bond | 0.47 | 2.00 | 0.08 | -6.68 | 10.00 | 0.10 | 0.12 | 1.00 |
| 5Y Bond | 0.46 | 1.38 | 0.10 | -5.80 | 10.61 | 0.07 | 0.19 | 0.90 |
| 2Y Bond | 0.42 | 0.77 | 0.13 | -3.69 | 8.42 | 0.08 | 0.37 | 0.76 |
| 1Y Bond | 0.40 | 0.50 | 0.16 | -1.72 | 5.61 | 0.08 | 0.59 | 0.62 |
| 1M TBill | 0.32 | 0.26 | | -0.00 | 1.52 | -0.05 | 1.00 | 0.12 |
| CPI | 0.31 | 0.45 | | -1.92 | 5.88 | -0.07 | 0.26 | -0.07 |
| **Monthly** | mean | std | Sharpe | min | max | correlation with | | |
| **1990-2014** | (%) | (%) | ratio | (%) | (%) | Stock | TBill | 10Y |
| Stock | 0.87 | 4.22 | 0.15 | -16.70 | 11.41 | 1.00 | 0.01 | -0.06 |
| 10Y Bond | 0.57 | 1.99 | 0.16 | -6.68 | 8.54 | -0.06 | 0.07 | 1.00 |
| 5Y Bond | 0.50 | 1.24 | 0.20 | -3.38 | 4.52 | -0.10 | 0.15 | 0.93 |
| 2Y Bond | 0.39 | 0.54 | 0.26 | -1.30 | 2.07 | -0.11 | 0.41 | 0.74 |
| 1Y Bond | 0.33 | 0.31 | 0.26 | -0.33 | 1.31 | -0.03 | 0.72 | 0.51 |
| 1M TBill | 0.25 | 0.19 | | -0.00 | 0.68 | 0.01 | 1.00 | 0.07 |
| CPI | 0.21 | 0.34 | | -1.92 | 1.22 | -0.04 | 0.18 | -0.16 |

Table 1 also reports the best and worst one-month returns for each of the securities. Not surprisingly, the stock market is the most risky with the largest range of minimum and maximum. During the sample period from 1942 to 2014, the worst one-month return was -21.58%, which happened in October 1987.

Also reported are the correlations between the stock returns and the bond returns. The correlation between these two markets is very weak and is also unstable. The correlation between stock and 10-year bond is 10% for the sample from 1942 through 2014 and -6% for the more recent sample from 1990 through 2014. Unlike the low correlation between stock and bond, the correlations between the bond returns are relatively high. The closer the maturity (e.g., 10Y and 5Y), the higher the correlation. We will come back and the investigate this issue in our next class when we do PCA (Principal Component Analysis) on bonds.

It is also interesting to see that the correlations between inflation (CPI) and the stock returns and 10Y are low and slightly negative. The correlation between inflation and

the 1M Tbill is about 26% for the entire sample and 18% for the more recent sample. Note that we are working with nominal interest rate, which is the sum of real interest rate and inflation. As you can see from Table 1, the average inflation is close to the 1-month Treasury bill, but slightly lower, implying that the real interest rate is on average positive.

- **The Cycle of Hot and Cold:** Using the average return of the one-month Treasury bill as the riskfree rate, we can calculate the Sharpe ratios of the equity and bond returns. From this perspective, bonds have been more attractive (higher average return and lower volatility) for the more recent sample period from 1990 through 2014.

  In fact, from the mid 1980s to today, the bond market condition has been quite favorable. The interest rates have been decreasing from the double digits in the early 1980s to today's near-zero. Some call it a 30-year bull market run. In addition to the favorable market condition, we have also seen the rise of MBS, junk bonds, OTC derivatives, asset-backed securities, all of which add to the business of fixed-income desks in investment banks.

  When Michael Lewis joined the training program in Salomon in 1985, the bond market was just getting hot, driven by the profitability in bonds. In 1986, other firms like Goldman Sachs were catching up with Salomon's bond expertise by hiring people away from Salomon (See, for example, *Money and Power* by Cohan). Within Salomon, as described in Michael Lewis' book, *Liar's Poker*, an entertaining (maybe too entertaining) book, the desired location was to be on a bond desk. Equity was looked down up, and "Equity in Dallas" was the equivalent of Siberia.

  But only ten years prior to that, bond was not at hot and equity was the place to go. Quoting Michael Lewis,

  > That, anyway, is what I was told. It was hard to prove any of it because the only evidence was oral. But consider the kickoff chuckle to a speech given to the Wharton School in March 1977 by Sidney Homer of Salomon Brothers, the leading bond analyst on Wall Street from the mid-1940s right through to the late 1970s. "I felt frustrated," said Homer about his job. "At cocktail parties lovely ladies would corner me and ask my opinion of the market, but alas, when they learned I was a bond man, they would quietly drift away."
  >
  > Or consider the very lack of evidence itself. There are 287 books about bonds in the New York Public Library, and most of them are about chemistry. The ones that aren't contain lots of ugly numbers and bear titles such as *All Quiet*

> *on the Bond Front*, and *Low-Risk Strategies for the Investor*. In other words, they aren't the sort of page turners that moisten your palms and glue you to your seat. People who believe themselves of social consequence tend to leave more of a paper trail, in the form of memoirs and anecdotiana. But while there are dozens of anecdotes and several memoirs from the stock markets, the bond markets are officially silent. Bond people pose the same problem to a cultural anthropologist as a nonliterate tribe deep in the Amazon.

By now, bond people are certainly not the equivalent of a nonliterate tribe deep in the Amazon. In fact, if you search Amazon for books on Finance, many of them were written by bond traders. So is this endless cycle of being hot and cold, in and out of favor. Whatever that can go up certainly has the potential to come down. The moment something is in favor marks the beginning of its decline.

Right now (Fall 2015), the interest rate is at a level as low as it can ever be, and the 30-year bull run in the bond market is approaching to an end. Most likely, the Fed will raise the Fed fund rate in its December FOMC meeting this year (Fall 2015). Inferring from the pricing in Fed fund futures, there is a 70% likelihood of a Fed hike at its December 15-16 meeting (Fall 2015). So we will know the result before our final exam on December 17 (Fall 2015).

In the mutual fund world, the famous bond fund, Pimco's Total Return, is a good representation of this cycle of bull and impending bear. As shown in Figure 1, the first observation of Pimco (Total Return Fund, Institutional Class) in my data was at the end of June 1987 with a total net asset value of $12.8 million. From 1987 to 2013, the fund, benefited from the favorable bond market condition, was in a steady ascend, reaching to its peak ($182.8 billion) in April 2013. This grow in the size of a mutual fund has two component: the market performance and fund flows. So the growth from $18 million to $182.8 billion was a combination of both. As we know, in the mutual fund word, flow chases performance. So the favorable condition in the bond market has a lot to do with the growth.

In recent months, the size of the fund has been decreasing quite rapidly. Figure 2 plots the total net asset value for all four classes of the fund. Of course, if you have been following the news since 2014, you would know that the internal powerful struggle and the clash of personalities also contributed to the fund outflow. But the clash of personality probably would not have escalated to such a degree had the bond market condition been favorable.

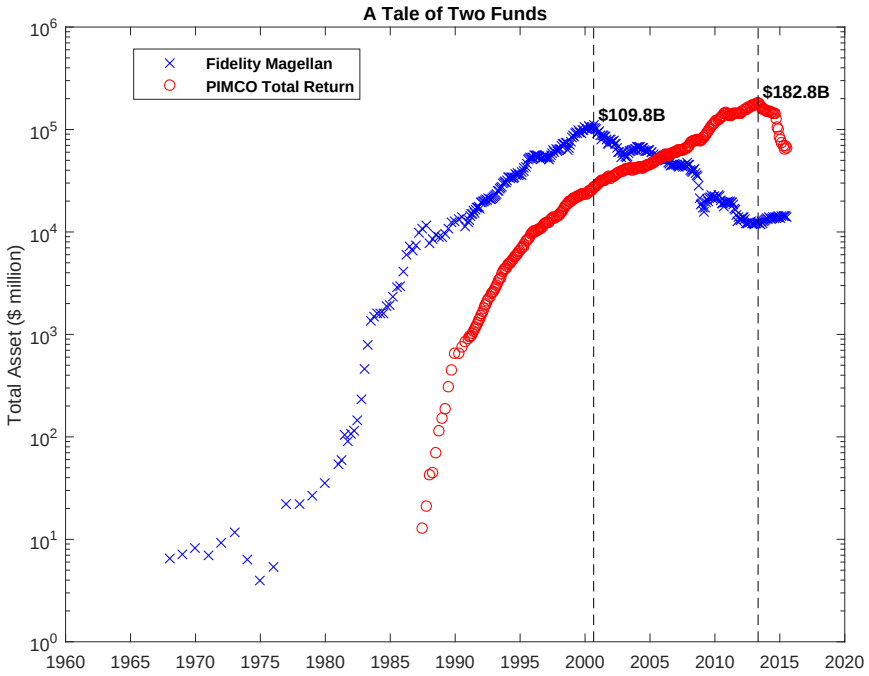


Figure 1: Total Net Asset Value, Fidelity Magellan and Pimco Total Return.

Also note that the plot is in log-scale, in an effort to damp the high growth rate. If it were plotted in a linear plot, the ups and downs would have been even more dramatic.

As another example of the force of the overall market condition versus the skills of an individual fund manager, I plot in Figure 1 the total net asset value of the once famous equity fund, Fidelity Magellan. The fund shows up in my data since May 1963 but the first reported total net asset value in my data was $6.5 million in December 1967. By December 1975, the fund was smaller at $5.4 million, most likely due to the bear market of 1973-74. In June 1976 Peter Lynch took over the fund. From 1976 to 1990, under Peter Lynch's management, the fund grew in size as well as in fame. After Peter Lynch's retirement in May 1990, the fund kept growing, thankful to the bull market of the late 1990s. The fund grew to its peak ($109.8 billion) in August 2000, and then started its decline after the Internet bubble burst. Right now (Fall 2015), it is a $14 billion fund, roughly the size when Peter Lynch retired from the fund in May 1990.

Cycles like those in Figure 1 are part and parcel of the financial markets. Such forces in financial markets should be humbling for any human being, no matter how successful this person might be. To attribute one's success entirely to one's talent is pure arrogance and ignorance. If you have not read the recent stories surrounding Bill Gross (the co-founder of Pimco), I would suggest that you do. At some point in your life,

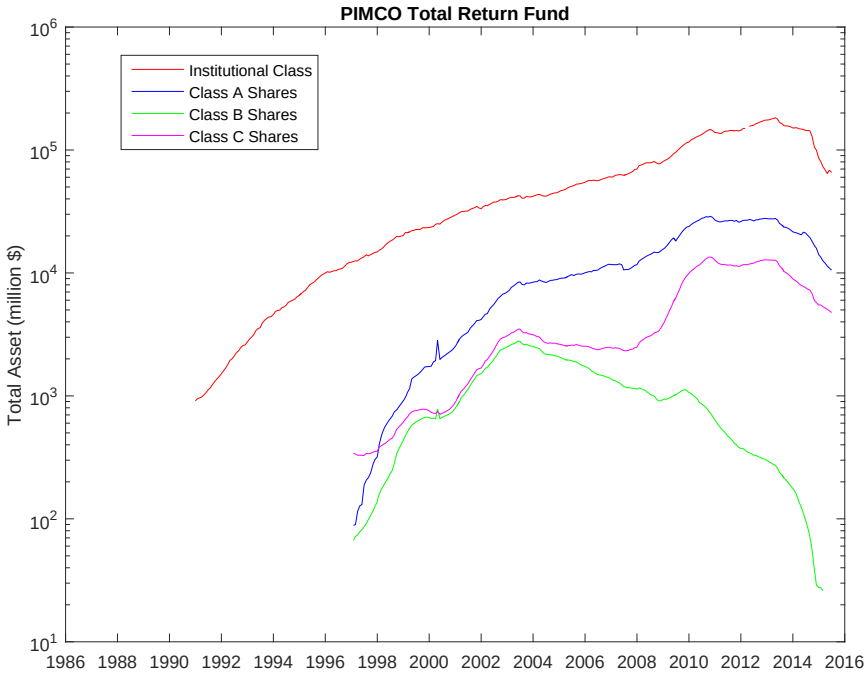


Figure 2: Total Net Asset Value, Pimco Total Return Fund.

you might get lucky and become successful. Try not to let your ego drive you too far. There are no worse enemies in your life than your own ego. In fact, your ego is you only enemy.

# 2 Bond Price and Yield: Duration and Convexity

- **Bond Price $P$ and Yield to Maturity $y$:** A Treasury yield curve involves Treasury bonds, notes, and bills. Treasury notes are issued in terms of 2, 3, 5, 7, and 10 years; Treasury bonds are issued at 30 years. A Treasury bond issued 25 years ago would have 5 years to maturity, same as a newly issued 5-year notes. But the coupon rates of the two bonds are different. Coupon bearing bonds are issued at par, making the coupon rate close to the yield to maturity at the time of issuance. Given the current low interest rate environment, the 30-year bond issued 25 years ago has a coupon rate that is higher than the newly issued 5 year notes. It is therefore a premium bond. There are also differences in liquidity, which we will talk about later.

  Throughout the fixed-income classes, I'll not make a distinction between notes and bonds and will refer to them simply as bonds. I'll use the notation of $P_t$ as the bond price at time $t$, and $y_t\%$ as the yield to maturity at time $t$. At issuance, a Treasury

bond is defined by the following parameters: **face value** = \$100; **coupon rate** = $c$; **maturity** = T years. These parameters are fixed throughout the life of the bond and will not change. Treasury bonds pay coupon semi-annually, and, at issuance, the coupon rate $c$ is chosen so that the bond is priced at par with $P = \$100$. As a result, the yield to maturity $y$ (semi-annual compounding) equals to the coupon rate $c$ when the bond was first issued.

Later, with the fluctuations in interest rates, both $P$ and $y$ will change. There is a deterministic relation between the two:

$$P = \sum_{n=1}^{2T} \frac{\frac{c}{2} \times 100}{\left(1 + \frac{y}{2}\right)^n} + \frac{100}{\left(1 + \frac{y}{2}\right)^{2T}}, \tag{1}$$

where both $c$ and $y$ are expressed in percentage. So an increasing interest rate environment after the issuance of the bond is bad news for long-only bond investors: $P$ decreases with increasing $y$ and the bond will be in discount ($P < \$100$). Conversely, a decreasing interest rate environment is good news such a long-only bond investor: $P$ increases with decreasing $y$ and the bond is in premium ($P > \$100$).

So Treasury bonds are not at all riskfree, and its volatility is driven by the volatility of the interest rate. Assuming the high credit quality of the US government, the Treasury bonds are considered to be almost default free. During the heat of the debt-ceiling crisis in 2011, the rating agency S&P downgraded the US Treasury from AAA to AA+. The financial markets were in a crisis mode and Treasury bonds actually appreciated in value because, out of the flight to quality, investors move their capital away from risky assets to ... the US Treasury bonds.

The relation between $P$ and $y$ as expressed in Equation (1) is a very important one, and we will come back to it again. For now, I would like you to keep the picture of Figure 3 in mind. This is what the payoff schedule of a bond looks like. Over the life of the bond, you collect small coupon payments every six months, and toward the end of the life of the bond, at maturity, you collect the last coupon payment plus the principal. You discount this cashflow by a constant interest rate $y$ using the discount function $1/(1 + y/2)^n$ for the $n$-th semi-annual payment. In doing this calculation, you link the bond price $P$ to its yield to maturity $y$. There is no uncertainty involved in this relationship. There is also no economics involved in this calculation. But the calculation becomes very handy as we move between $P$ and $y$. Concepts such as duration and convexity arise out of this calculation.

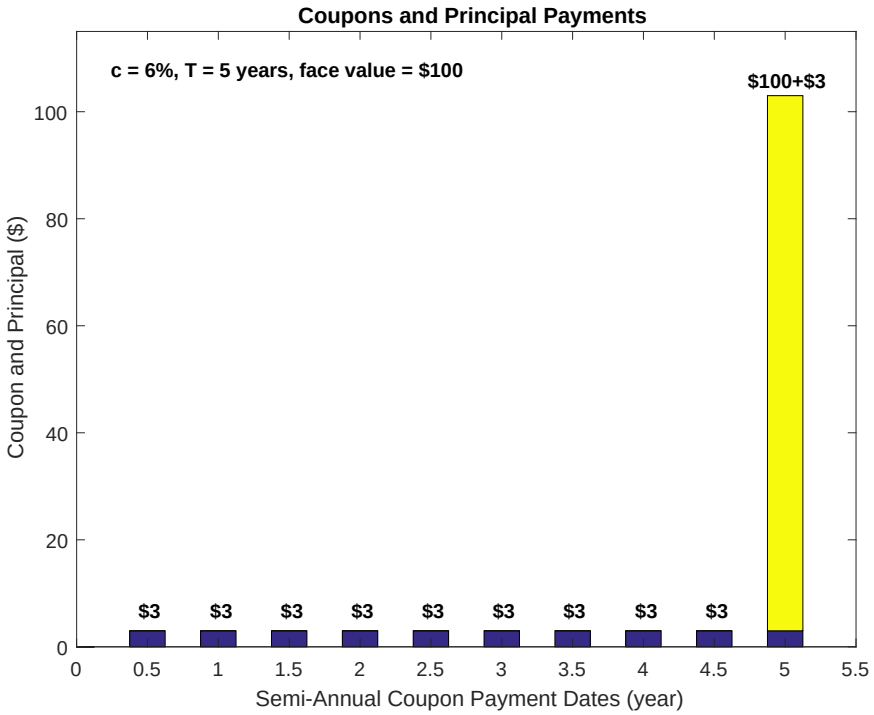


Figure 3: Coupon and Principal Payment Dates

- **Treasury Yield Curve:** As shown in Figure 4, a Treasury yield curve is plot of yield against maturity, for Treasury bonds of varying maturities. Treasury bonds are traded in terms of market prices $P$. So a yield curve is constructed using the market prices of individual Treasury bonds. In Figure 4, the green dots are Treasury bills, the blue dots are Treasury notes, and the purple dots are old Treasury bonds. For example, the yield curve in Figure 4 was plotted for November 8, 1994. For a purple dot with a maturity of seven years, the bond was issued 23 years ago in 1971 as a 30-year Treasury bond.

  As you can see, the yield curve is not created in vacuum. It is made up of individual bonds. In fact, the creation of a yield curve is not a simple task. The various bonds have different liquidity: the old bonds are typically less liquid while the new bonds/notes are typically very liquid. The liquidity effect shows up in the market prices of these bonds: illiquid bonds are cheaper than the liquid bonds. As a result, in constructing the yield curve, considerations such as liquidity take place. I do not want to make you a specialist in curve fitting, but if we have time in the next class, I will talk more about curve fitting.

  Focusing back on the yield curve in Figure 4, we see that on this day, the term structure is upward sloping. The short end of the yield curve is about 4.6%, the 2-year yield

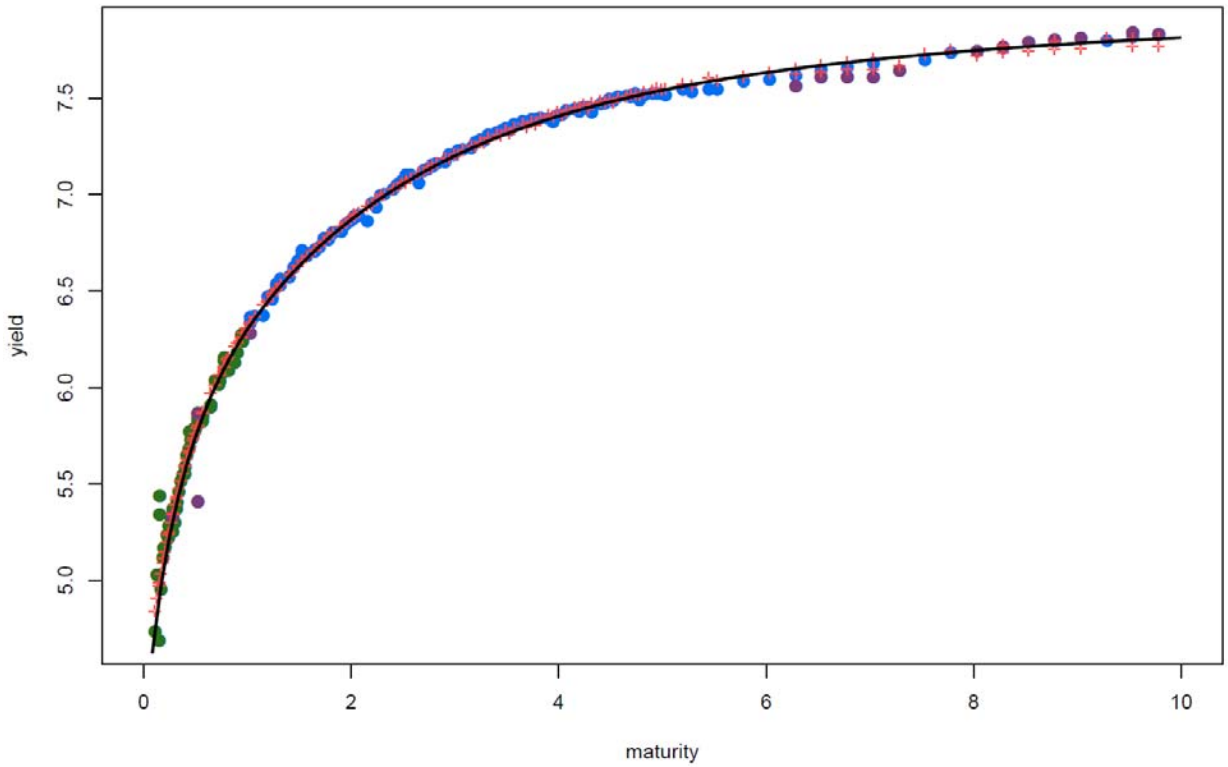


Figure 4: Treasury Yield Curve on November 8, 1994.

is about 6.8%, and the 10-year yield is at 7.8%. This makes the 10y to 2y spread at about 100 basis points. For bonds of similar maturities, the spreads are quite tight, indicating active arbitrage activities on the yield curve. By comparison, the yield curve on December 11, 2008, plotted in Figure 5, looks quite dramatic. Bonds are very similar maturities are trading at a yield spread in the order of 50 basis points. During normal market conditions, spreads so wide would never happen in this market. Of course, December 2008 was not normal. This picture indicates the lack of arbitrage activities in 2008, even in the most liquid market.

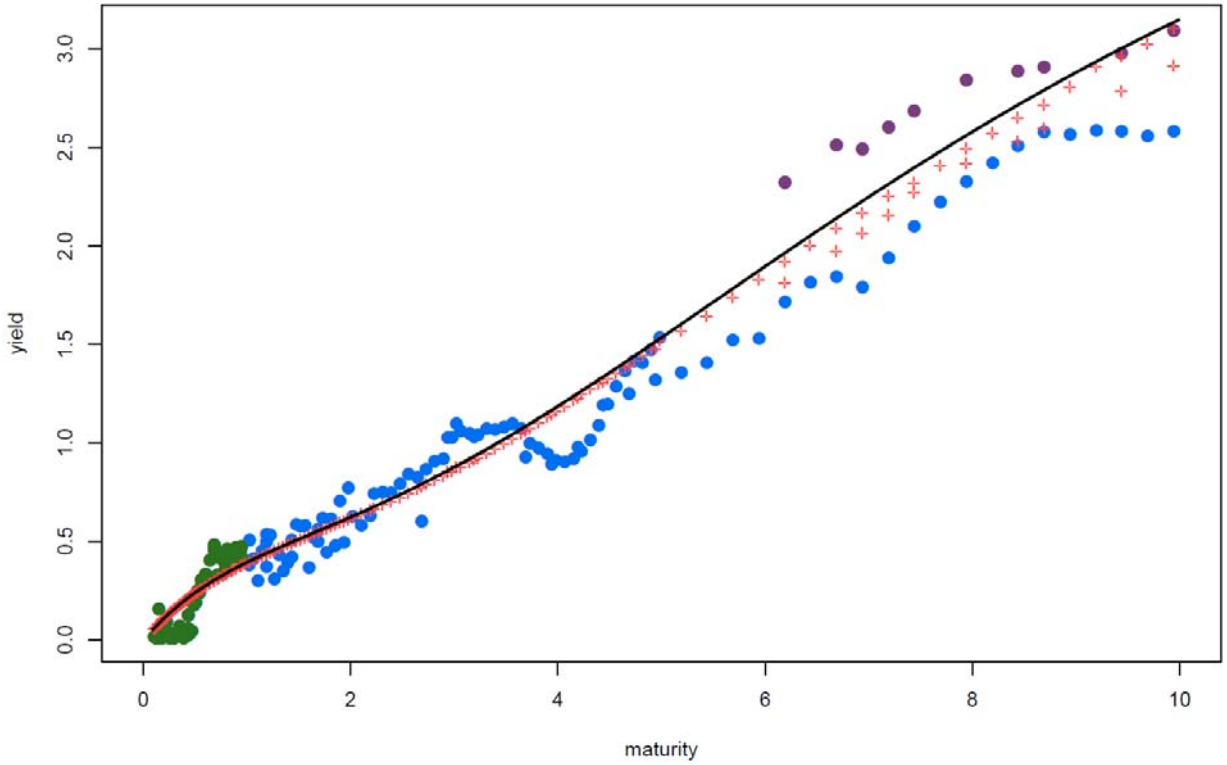


Figure 5: Treasury Yield Curve on December 11, 2008.

- **Time-Varying Yields:** To understand how the yield curve move over time, Figure 6 plots the time-series of Treasury constant maturity yields for a few selected maturities.

  These constant maturity yields are calculated daily by using market prices of Treasury bonds as the input. And the output is the par-coupon yields of varying maturities. Effectively, these are interpolated yields for the a set of fixed maturity of interest (e.g., 1, 2, 3, 5, 7, 10, 20, and 30 years). Again, to know what is really going on, we need to

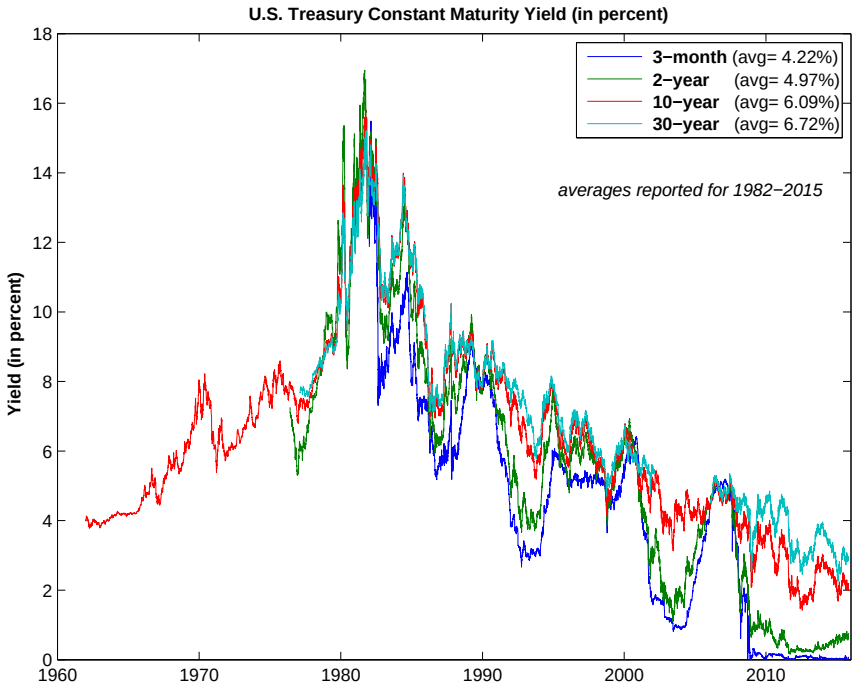


Figure 6: Time-Series of Treasury Constant Maturity Yields.

spend some time on curve fitting. For those who are interested, this is a not so useful explanation from the Treasury department, but it is better than nothing.

Let's now used these CMT yields and see how the yield curve vary over time. As shown in Figure 6, most of the time, the yield curve is upward sloping. Using data from 1982 to today, the 2-year CMT yield is on average 4.97%, the 10-year yield is on average 6.09%, and the 30-year yield is on average 6.72%. So the spread of 10y to 2y is on average 100 basis points. There are also times when the yield curve is not so steep or even inverting. We will take a closer look later on these events. Also notice that the green line (2yr yield) is picking up in recent days. The 2yr yield is a policy sensitive yield and is moving up in anticipation of a rate hike.

Also notice the missing 30yr yield in Figure 6 from early 2002 to early 2006. In late 2001, facing projections of burgeoning surpluses, the Treasury decided to stop issuing the 30-year bond to save tax payers money. In late 2005, the Treasury decided to re-introduce the 30-year bond and held its first auction in fives years on February 9, 2006.

Using these CMT yields, let's also calculate the daily volatility of the Treasury yields. As shown in Table 2, using daily data from 1982 to today, the standard deviation of

the daily changes in the 3M Tbill rate is about 7.63 basis points. The 2Y and 10Y yields are slightly less volatile, at around 6.8 basis points. In recent years, however, the volatility is low for the short end because of the monetary policy. In general, however, the short end of the yield is typically more volatile, although the different in volatility is not huge. In other words, when measured in the yield space, the volatility across different maturity is comparable. But when it comes to the return space, the volatility across different maturity will be very different because of the difference in duration, which we will see shortly.

Table 2: Summary Statistics of Daily Changes in Treasury Yields

| sample | maturity | std (bp) | min (bp) | date | max (bp) | date |
|---|---|---|---|---|---|---|
| 1982-2015 | 3M | 7.63 | -104 | 19820222 | 169 | 19820201 |
| | 2Y | 6.86 | -84 | 19871020 | 80 | 19820201 |
| | 10Y | 6.80 | -75 | 19871020 | 44 | 19820201 |
| | 30Y | 6.30 | -76 | 19871020 | 42 | 19820201 |
| 1990-2008 | 3M | 5.18 | -64 | 20070820 | 58 | 20001226 |
| | 2Y | 6.05 | -54 | 20010913 | 36 | 19940404 |
| | 10Y | 5.78 | -23 | 19950613 | 39 | 19940404 |
| | 30Y | 4.99 | -33 | 20011031 | 32 | 19940404 |
| 2008-2015 | 3M | 4.94 | -81 | 20080917 | 76 | 20080919 |
| | 2Y | 4.86 | -45 | 20080915 | 38 | 20080919 |
| | 10Y | 6.42 | -51 | 20090318 | 24 | 20080930 |
| | 30Y | 6.12 | -32 | 20081120 | 28 | 20110811 |

Table 2 also reports the largest one day movements for these yields. Let me link a few of these extreme movements in yield to the events at the time:

- October 20, 1987 was the day after the 1987 stock market crash.
- April 1994 was a very testy time in the bond market because of monetary policy tightening by Chairman Greenspan.
- September 15 to 19, 2008 was the week of Lehman default and AIG bailout. TBill rates first decreased sharply (increased in value) because of flight to quality and then bounced back on September 19.
- On March 18, 2009, the Fed made the following announcements, which were summarized in Chairman Ben Bernanke's recent book. *The overall package was designed to get markets' attention, and it did. We announced that we planned to*

*increase our 2009 purchases of mortgage-backed securities guaranteed by Fannie, Freddie, and Ginnie Mae to $1.25 trillion, an increase of $750 billion. We also doubled, from $100 billion to $200 billion, our planned purchases of the debt issued by Fannie and Freddie to finance their own holdings. We would also buy $300 billion of Treasuries over the next six months, our first foray into Treasury purchases. Finally, we strengthened our guidance about our plans for our benchmark interest rate, the federal funds rate. In January, we had said that we expected the funds rate to be at exceptionally low levels "for some time." In March, "for some time" became "for an extended period." We hoped that this new signal on short-term rates would help bring down long-term rates.*

  - The across-the-board increase in yield on February 1, 1982 was likely caused by the monetary policy tightening under Chairman Paul Volcker.

Overall, the numbers presented in Table 2 give us a baseline in observing and judging the daily movements in interest rates. A one-sigma move in this market is about 6 to 7 basis points. A daily movement of 25 basis points is unusual for this market.

- **Dollar Duration:** There are two measures of duration that is important for us to know. The dollar duration is defined as

$$-\frac{\partial P}{\partial y} = \frac{1}{1+\frac{y}{2}}\left[\sum_{n=1}^{2T} \frac{n}{2} \times \frac{\frac{c}{2}\times 100}{\left(1+\frac{y}{2}\right)^n} + T \times \frac{100}{\left(1+\frac{y}{2}\right)^{2T}}\right], \tag{2}$$

which is the negative of dollar change in bond price per unit change in yield. Given that a typical change in yield is measured in basis points, the often used DV01 measure scales the dollar measure by 10,000:

$$\text{DV01} = \text{Dollar Duration}/10,000\,,$$

which measures the negative change in bond price per one basis point change in yield.

Figure 7 plots the bond price $P$ as a function of yield $y$ for a ten-year bond with coupon rate of 6%. Effectively, it plots the relation between $P$ and $y$ in Equation (1). As we can see, $P$ is inversely related to $y$: decreasing $y$ is coupled with increasing $P$. Also, the relation is not linear. But if we would like to approximate the relation linearly, we can pick a level of $y$, say $y = 6\%$ and $P = \$100$ and draw a tangent line at that point. As you've been taught many times in the past, the slope is $\partial P/\partial y$ as calculated in Equation (2). In other words, the dollar duration is the negative of the slope.

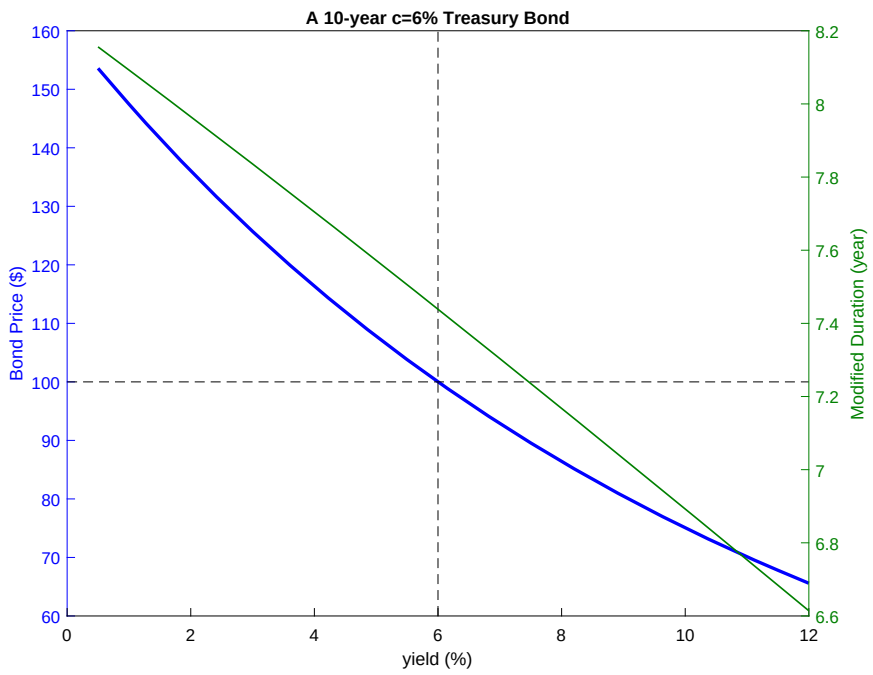


Figure 7: Bond Price as a Function of Yield and Duration as a Function of Yield

So if I would like to know how much I will lose when the ten-year Treasury yield suddenly increases by 10 basis points, I can use the linear approximation:

$$\Delta P_t = P_t - P_{t-1} \approx -\mathrm{D}^{\$} \times (y_t - y_{t-1}) = -\mathrm{D}^{\$} \times \Delta y_t = -D^{\$} \times \frac{10}{10,000} = -\mathrm{DV01} \times 10\,\mathrm{bps}$$

Going back to Figure 7, let's still focus just on the blue line. We notice that when $y$ decreases, the slope gets steeper; when $y$ increases, the slope gets flatter. This is because the relation between $P$ and $y$ as defined by Equation (1) is convex. For an investor holding a long position in bond, he would very much welcome this feature: profits due to decreasing $y$ are amplified and losses due to increasing $y$ are dampened.

- **Modified Duration:** The modified duration is defined as

$$-\frac{1}{P}\frac{\partial P}{\partial y} = \frac{1}{1+\frac{y}{2}} \frac{\sum_{n=1}^{2T} \frac{n}{2} \times \frac{\frac{c}{2}\times 100}{\left(1+\frac{y}{2}\right)^n} + T \times \frac{100}{\left(1+\frac{y}{2}\right)^{2T}}}{\sum_{n=1}^{2T} \frac{\frac{c}{2}\times 100}{\left(1+\frac{y}{2}\right)^n} + \frac{100}{\left(1+\frac{y}{2}\right)^{2T}}} \tag{3}$$

It is the dollar duration divided by the bond price. So its focus is on the profit/loss as

a fraction of the position:

$$R_t = \frac{\Delta P_t}{P_{t-1}} = \frac{P_t - P_{t-1}}{P_{t-1}} - \approx \mathrm{D}^{\mathrm{mod}} \times (y_t - y_{t-1}) = -\mathrm{D}^{\mathrm{mod}} \times \Delta y_t$$

Dollar durations and modified durations are used for different purposes. If we are interested in the profit/loss in dollar terms, we go with the dollar duration, but if we interested in the profit/loss in the return space, we go with the modified duration.

As shown in Equation (3), the modified duration is a normalized measure and the unit is in year. In dealing with coupon bonds, it is always useful to go to the extreme and think first in terms of zero-coupon bond. For a $T$-year zero-coupon bond, the modified duration is $T$ divided by $(1 + y/2)$. If instead of semi-annual compounding, the yield $y$ is continuously compounded, then the modified duration of a $T$-year zero-coupon bond is simply $T$.

For a bond with semi-annual coupon payments, the modified duration is a weighted sum of all of the coupon payment dates, 0.5, 1.0, 1.5, ..., and $T$ years. Except for the final date $T$, the $n$-th coupon dates are weighted by $\frac{c/2 \times 100}{(1+y/2)^n}$. The last date $T$ carries a disproportionately high weight because of the principal payment \$100. Because of this, the weighting is always tilted toward the final date $T$. To be more precise, date $T$ is weighted by $\frac{c/2 \times 100 + 100}{(1+y/2)^{2T}}$. For a coupon rate of 6%, $c/2 \times 100 + 100$ is 103, easily overpowering $c/2 = 3$.

You might wonder what happens when we have a really aggressive discount rate $y$, say $y = 10\%$? Well, let's consider the two extreme points: $\frac{1}{(1+y/2)^n}$ for the first coupon payment $n = 1$ and $\frac{1}{(1+y/2)^{2T}}$ for the final date $T$. Plugging $y = 10\%$, we have $\frac{1}{(1+y/2)} = 0.9524$ and $\frac{1}{(1+y/2)^{2T}} = 0.3769$ for $T = 10$. As you can see, even with this very aggressive discount rate discounting over a 10-year period, the principal payment of \$100 still dominates the calculation.

This is why, as you can see in Table 3, the modified duration of a ten-year bond is close to 10, especially when $y$ is low. As $y$ gets higher, this discounting effect becomes relatively more important, pushing the "center of gravity" away from $T$. As a result, the modified duration gets smaller.

Building on this analogy of "center of gravity" a little bit more, let's go back to the picture in Figure 3, which is a useful picture to have in our head when doing bond math. At least this is how I do the math. I imagine that there is a center of gravity along the horizontal dimension. Its gets pulled/pushed left and right, depending on the

Table 3: Modified Duration

| yield $y$ | 2% | 3% | 4% | 5% | 6% | 6% | 6% | 7% | 10% |
|---|---|---|---|---|---|---|---|---|---|
| coupon $c$ | 2% | 3% | 4% | 5% | 4.8% | 6% | 7.2% | 7% | 10% |
| $T = 1$ | 0.99 | 0.98 | 0.97 | 0.96 | 0.96 | 0.96 | 0.95 | 0.95 | 0.93 |
| $T = 2$ | 1.95 | 1.93 | 1.90 | 1.88 | 1.87 | 1.86 | 1.84 | 1.84 | 1.77 |
| $T = 3$ | 2.90 | 2.85 | 2.80 | 2.75 | 2.74 | 2.71 | 2.68 | 2.66 | 2.54 |
| $T = 5$ | 4.74 | 4.61 | 4.49 | 4.38 | 4.36 | 4.27 | 4.18 | 4.16 | 3.86 |
| $T = 7$ | 6.50 | 6.27 | 6.05 | 5.85 | 5.81 | 5.65 | 5.51 | 5.46 | 4.95 |
| $T = 10$ | 9.02 | 8.58 | 8.18 | 7.79 | 7.71 | 7.44 | 7.21 | 7.11 | 6.23 |
| $T = 20$ | 16.42 | 14.96 | 13.68 | 12.55 | 12.12 | 11.56 | 11.13 | 10.68 | 8.58 |
| $T = 30$ | 22.48 | 19.69 | 17.38 | 15.45 | 14.46 | 13.84 | 13.39 | 12.47 | 9.46 |

relative weights between the last date $T$ and the other coupon dates. Getting pushed to the left results in a smaller duration and getting pull to the right results in a larger duration.

For example, consider two bonds with the same $y$ and same $T$ but different coupon rate $c$. It could be that one bond was issued back in 1990 as a 30-year bond and has five year to maturity. The other bond is a newly issued 5-year notes. Assuming a flat term structure of interest rate, the yields of these two bonds are the same, but their coupon rates are different (so are their bond prices). Which one has a higher duration? The one with lower $c$ has its center of gravity closer to $T$. As a result, it has a higher duration.

Generally, it is useful to have a table like that in Table 3 handy, or build a function in Excel to calculate the modified duration of a bond for give coupon $c$, yield $y$, and maturity $T$. Historically, the average 10-year yield is about 6%. It is useful to know that, for a 10-year par coupon bond with $c = 6\%$, its modified duration is around 7.44 years. (Not precisely 7.44, but a number around 7 or 8.) In recent years, interest rates have been low, implying a relatively high duration for bonds. Right now (Fall 2015), the 10-year yield is at 2.34%. It would be useful to know that a 10-year par coupon bond with $c = 2\%$ has a modified duration around 9 years. The current 5-year yield is at 1.72%, and it is useful to know that a 5-year par coupon bond with $c = 2\%$ has a modified duration around 4.75 years. There is no need to memorize these numbers, but to have a rough sense in terms of orders of magnitudes would be handy.

For example, we know that a typical one-day one-sigma move in 10-year yield is about 6.8 basis points. How much does that translate to return volatility? Recall that,

$R_t \approx \mathrm{D}^{\mathrm{mod}} \times \Delta y_t$. So, $\mathrm{std}(R_t) \approx \mathrm{D}^{\mathrm{mod}} \times \mathrm{std}\,(\Delta y_t)$. For a 10-year bond with a duration of 7.44, a 6.8-bps volatility in $\Delta y_t$ translates to $6.8 \times 7.44 = 50.6$ basis points in daily return volatility. Right now (Fall 2015), in a low interest rate environment, duration is high. For the same amount of volatility in $\Delta y_t$, the bond return volatility would be higher because of the higher duration.

As another example, suppose you believe that the 30-year bond is priced cheap relative to the yield curve. Your model tells you that the spread between the 30-year bond and the curve (generated by your model) is about 10 basis points. You believe that this spread is due to temporary illiquidity in 30-year bonds and will converge to close to zero later on. How much does this 10 basis points translate to return? Right now (Fall 2015), the 30-year yield is at 3.12%. Table 3 tells us that at this rate, the modified duration is about 20 years. So $R_t \approx -\mathrm{D}^{\mathrm{mod}} \times \Delta y_t = -20 \times (-10\,\mathrm{bps}) = 2\%$.

- **Duration and Convexity:** Concepts such as duration and convexity are only meaning because we work in the yield space and the profit/loss is in the dollar space. As such, duration serves as a bridge that connects the bond price to yield:

  – Dollar Duration:

$$\Delta P_t = P_t - P_{t-1} \approx -\mathrm{D}^{\$} \times (y_t - y_{t-1}) = -\mathrm{D}^{\$} \times \Delta y_t$$

  – Modified Duration:

$$R_t = \frac{\Delta P_t}{P_{t-1}} = \frac{P_t - P_{t-1}}{P_{t-1}} - \approx \mathrm{D}^{\mathrm{mod}} \times (y_t - y_{t-1}) = -\mathrm{D}^{\mathrm{mod}} \times \Delta y_t$$

In addition to this linear approximation through duration, we also notice that the relation between price and yield is not linear but convex. So convexity is introduced as a second-order approximation to improve upon the first order, linear approximation. In this class, we will not go for the exact formula for this second order approximation. If one day, you become a bond trader/portfolio manager, than you might be busy with convexity hedging. Even then, you might notice that the term structure of interest rate is not flat, which could cause quite a bit of problem for your first order duration hedge.

Let me close by talking about one intuition associated with convexity that is important. The relation between duration and yield is as plotted in Figure 7. With decreasing $y$, duration increases. As a result, the profit from holding a bond gets amplified. This

effect is not symmetric in losses because with increasing $y$, duration decreases. As a result, the loss associated with holding a bond gets dampened. This positive convexity makes bond more attractive than a security that is linear in $y$. Later on, we will see a fixed-income security (Mortgage-Backed Securities) with negative convexity and use bonds (with positive convexity) to do duration hedge.

# 3 The Universe of Fixed Income Securities

Fixed-income securities share one thing: exposures to the Treasury yield curve. Most of these securities have an added component of credit risk. Muni's are bonds issued by municipalities, whose default probability is higher than the US government. The recent bankruptcy of Detroit is one example. Corporate bonds are issued by individual corporations, which also include credit risk. Agency bonds are issued by the government sponsored agencies (GSE) like Fannie and Freddie. After the government takeover in 2008, these bonds are explicitly backed by the US government. Prior to the takeover, it was implicitly backed by the government. For most of the fixed-income securities, the Treasury yield curve serves as a benchmark. Credit-sensitive instruments such as corporate bonds are usually quoted in terms of its spread relative to the US treasury yield.

Table 4 gives a summary of the US bond market. It gives us a sense of the relative size of the various components of the fixed-income market. In later classes, we will study the corporate bond market and will also touch upon the mortgage backed securities.

Table 4: Outstanding US Bond Market Debt in $ Billions

| | Muni | Treasury | Mortgage Related | Corp Debt | Agency Bonds | Money Markets | Asset Backed | Total |
|---|---|---|---|---|---|---|---|---|
| 1980 | 399.4 | 623.2 | 111.4 | 458.6 | 164.3 | 480.7 | | 2,237.7 |
| 1981 | 443.7 | 720.3 | 127.0 | 489.2 | 194.5 | 593.7 | | 2,568.4 |
| 1982 | 508.0 | 881.5 | 177.1 | 534.7 | 208.8 | 622.7 | | 2,932.8 |
| 1983 | 575.1 | 1,050.9 | 248.3 | 575.3 | 209.3 | 638.3 | | 3,297.2 |
| 1984 | 650.6 | 1,247.4 | 302.9 | 651.9 | 240.4 | 777.1 | | 3,870.4 |
| 1985 | 859.5 | 1,437.7 | 399.9 | 776.6 | 261.0 | 950.9 | 1.2 | 4,686.7 |
| 1986 | 920.4 | 1,619.0 | 614.7 | 959.3 | 276.6 | 998.6 | 11.3 | 5,399.8 |
| 1987 | 1,012.0 | 1,724.7 | 816.0 | 1,074.9 | 308.3 | 1,125.8 | 18.1 | 6,079.7 |
| 1988 | 1,080.0 | 1,821.3 | 973.6 | 1,195.8 | 370.7 | 1,263.0 | 25.8 | 6,730.1 |
| 1989 | 1,129.8 | 1,945.4 | 1,192.7 | 1,292.4 | 397.5 | 1,359.5 | 37.3 | 7,354.6 |
| 1990 | 1,178.6 | 2,195.8 | 1,340.1 | 1,350.3 | 421.5 | 1,328.9 | 66.2 | 7,881.5 |
| 1991 | 1,272.1 | 2,471.6 | 1,577.1 | 1,454.6 | 421.5 | 1,215.7 | 91.7 | 8,504.3 |
| 1992 | 1,295.4 | 2,754.1 | 1,774.3 | 1,557.1 | 462.4 | 1,157.9 | 116.4 | 9,117.6 |
| 1993 | 1,361.7 | 2,989.5 | 2,209.0 | 1,782.8 | 550.8 | 1,143.6 | 132.5 | 10,170.0 |
| 1994 | 1,325.8 | 3,126.0 | 2,352.9 | 1,931.1 | 727.7 | 1,229.1 | 161.9 | 10,854.5 |
| 1995 | 1,268.2 | 3,307.2 | 2,432.1 | 2,087.5 | 924.0 | 1,367.6 | 214.9 | 11,601.4 |
| 1996 | 1,261.6 | 3,459.7 | 2,606.4 | 2,247.9 | 925.8 | 1,572.7 | 296.8 | 12,371.0 |
| 1997 | 1,318.5 | 3,456.8 | 2,871.8 | 2,457.5 | 1,021.8 | 1,871.1 | 392.5 | 13,390.0 |
| 1998 | 1,402.7 | 3,355.5 | 3,243.4 | 2,779.4 | 1,302.1 | 2,091.9 | 477.8 | 14,652.8 |
| 1999 | 1,457.1 | 3,266.0 | 3,832.2 | 3,120.0 | 1,620.0 | 2,452.7 | 583.5 | 16,331.5 |
| 2000 | 1,480.7 | 2,951.9 | 4,119.3 | 3,400.5 | 1,853.7 | 2,815.8 | 699.5 | 17,321.5 |
| 2001 | 1,603.4 | 2,967.5 | 4,711.0 | 3,824.6 | 2,157.4 | 2,715.0 | 811.9 | 18,790.8 |
| 2002 | 1,762.8 | 3,204.9 | 5,286.3 | 4,035.5 | 2,377.7 | 2,637.2 | 902.0 | 20,206.3 |
| 2003 | 1,900.4 | 3,574.9 | 5,708.0 | 4,310.4 | 2,626.2 | 2,616.1 | 992.7 | 21,728.6 |
| 2004 | 2,821.2 | 3,943.6 | 6,289.1 | 4,537.9 | 2,700.6 | 2,996.1 | 1,096.6 | 24,385.1 |
| 2005 | 3,019.3 | 4,165.9 | 7,206.4 | 4,604.0 | 2,616.0 | 3,536.6 | 1,275.0 | 26,423.2 |
| 2006 | 3,189.3 | 4,322.9 | 8,376.0 | 4,842.5 | 2,634.0 | 4,140.0 | 1,642.7 | 29,147.3 |
| 2007 | 3,424.8 | 4,516.7 | 9,372.6 | 5,254.3 | 2,906.2 | 4,310.8 | 1,938.8 | 31,724.2 |
| 2008 | 3,517.2 | 5,783.6 | 9,457.6 | 5,417.5 | 3,210.6 | 3,939.3 | 1,799.3 | 33,125.2 |
| 2009 | 3,672.5 | 7,260.6 | 9,341.6 | 5,934.5 | 2,727.5 | 3,243.9 | 1,682.1 | 33,862.7 |
| 2010 | 3,772.1 | 8,853.0 | 9,221.4 | 6,543.4 | 2,538.8 | 2,980.8 | 1,476.3 | 35,385.9 |
| 2011 | 3,719.4 | 9,928.4 | 9,043.8 | 6,618.1 | 2,326.9 | 2,719.3 | 1,330.0 | 35,685.9 |
| 2012 | 3,714.4 | 11,046.1 | 8,814.9 | 7,049.6 | 2,095.8 | 2,612.3 | 1,253.6 | 36,586.7 |
| 2013 | 3,671.2 | 11,854.4 | 8,720.1 | 7,458.6 | 2,056.9 | 2,713.7 | 1,252.5 | 37,727.3 |
| 2014 | 3,652.4 | 12,504.8 | 8,729.4 | 7,846.2 | 2,028.7 | 2,903.3 | 1,336.5 | 39,001.3 |
| 2015Q1 | 3,694.0 | 12,630.2 | 8,688.9 | 7,965.1 | 1,975.6 | 2,879.2 | 1,361.3 | 39,194.4 |

Jun Pan | MIT Sloan School of Management | 253-3083
junpan@mit.edu | 15.433-15.4331 Financial Markets, Fall 2015 | E62-624

# Class 22: Fixed Income, Term Structure Models

This Version: November 23, 2016[1]

## 1 Term Structure Models

- **The Challenge from the Data:** In the fixed income market, term structure models are used to model interest rates. The challenge from the data has two dimensions. First, it should take into account of how the interest rates move over time. Second, for a given time, it should be able to model the yield curve, also called the term structure of interest rates. Figure 1 is a good summary of these two challenges from the data: a good term structure model should be able to capture the dynamic variations of the level of interest rates and the shape of the yield curve.

  These two demands from the data are very similar to those in the equity market. A good model for stock market returns should be able to take into account of how stock returns vary over time, as well as how, for a give time, the cross-section of stocks are priced in relation to one another. In the equity market, an i.i.d. model for stock returns is a reasonable approximation. As such, the dynamics for stock returns are really simple: constant expected return $\mu$, constant volatility $\sigma$, and unpredictable random shocks $\epsilon_{t+1}$. Cross-sectionally, the expected stock returns are linked to one another through their exposures (i.e., betas) to risk factors in a model such as the CAPM. As such, the CAPM model is a static model with constant expected returns and constant beta.

  The need for a dynamic model shows up when we investigated the time-varying volatility in our volatility class and stochastic volatility in our options class. Here in this class, we have a chance to take a closer look at these dynamic models.

- **Term Structure Models, Historical Development:** Term structure models were developed in the mid-1970s by Cox, Ingersoll and Ross (1985) and Vasicek (1997). You

[1] A small correction of Figure 3, which was missing a portion of the Fed target rate.

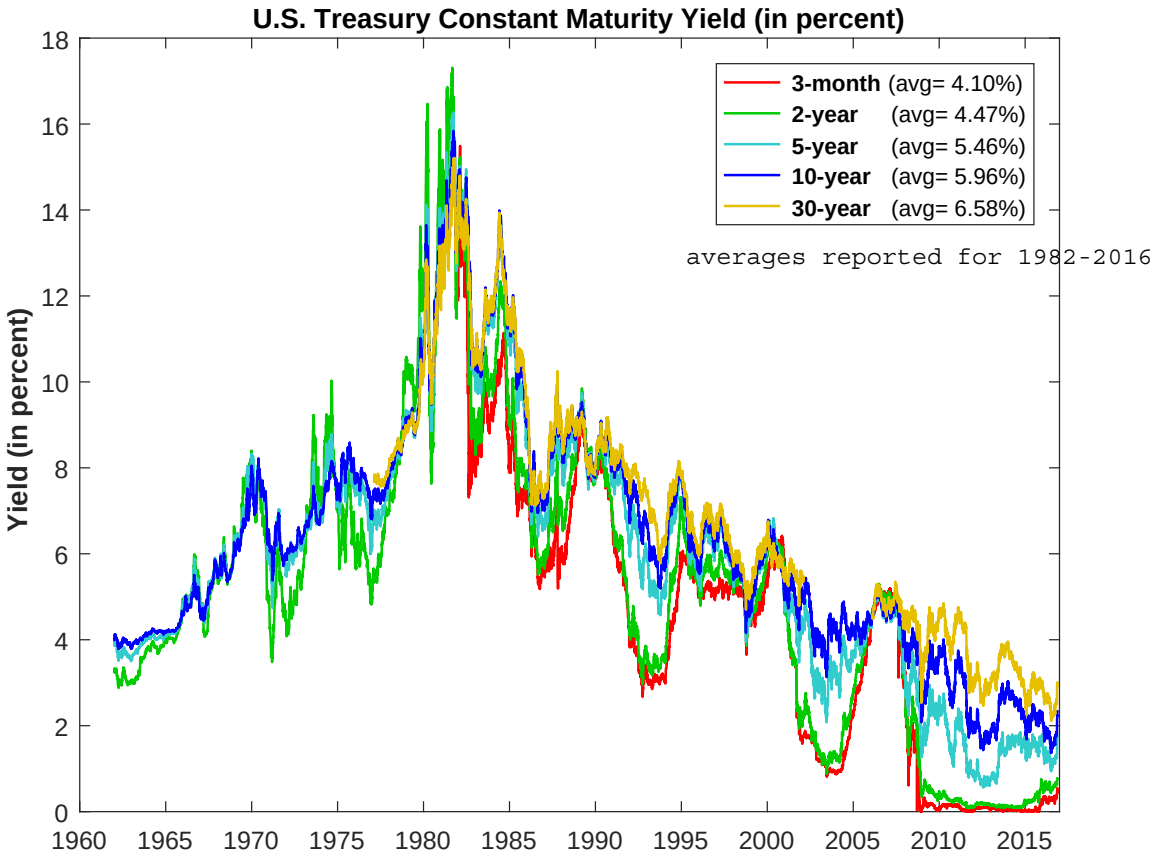


Figure 1: Time-Series of Treasury Constant Maturity Yields.

might notice that the CIR paper was published in 1985, but it was really a product of the mid-1970s. These term structure models were a continuation of the work done by Black, Merton, and Scholes, who popularized the application of continuous-time models in Finance. Like the Black-Scholes model before them, these term-structure models use the stochastic processes studied by mathematicians and physicists. For example, the CIR model builds on the Feller process and the Vasicek model builds on the Ornstein-Uhlenbeck process. In both cases, the starting point is the instantaneous short-rate $r_t$, which is modeled by a stochastic process (OU or Feller). The entire yield curve is priced using the dynamics of this one short rate. As such, the CIR and Vasicek models are one-factor short-rate models.

The second wave of term structure models came in the 1990s. When I entered the Stanford PhD program in 1995, I was just in time to catch the excitement surrounding term structure models. Relative to the original models of CIR and Vasicek, the effort of the new generation of term structure models is to be empirically relevant. From the work of Litterman and Scheinkman (1991), it became clear that a one-factor model will not be able to capture the entire shape of the yield curve. Unlike the stock market, where you can dismiss the risk uncaptured by the model as idiosyncratic risk, we do not have the luxury of dismissing common risk factors in the fixed income market (e.g., the slope factor).

These multifactor models quickly found their way into the "real" world. It is my understanding that each investment bank has its own proprietary term-structure model. And I was told by some practitioners that the industry has the best and most sophisticated term structure models. And they use these models to manage and hedge interest-rate risk (level, slope, convexity, volatility, etc) as well as to price interest-rate derivatives and other rate-sensitive instruments such as MBS. Looking back, I can now understand why during the mid-1990s, the Wall Street hired so many physicists and mathematicians. Most of my classmates in Physics ended up on Wall Street. I can also understand the sudden demand for more sophisticated term structure models in the 1990s. The fixed income desks were very profitable and the range and trading volume of their fixed income products were also expanding very rapidly during that time.

By now, the excitement surrounding term structure models has all but fizzled out. As a PhD student at Stanford, I spent much more time learning and working on term-structure models than anything else I did there. Since coming to Sloan in 2000, I have not made much use of that part of my training. Nevertheless, I am very grateful to my advisers at Stanford for having trained me in this area. As I wrote earlier in my

lecture notes, not everything we do in life is of practical use. Still, they are useful and meaningful in our growth process.

For our class, however, I don't want to emphasize too much on the modeling part, because it takes quite a bit of mathematical skills. Instead, I would like to use the term structure models as a way for us to understand conceptually how the various parts of the yield curve are connected through a pricing model and the role of the risk factors in generating the pricing results.

- **Bond Pricing in Continuous-Time:** Let $r_t$ be the time-$t$ instantaneous short rate. Let today be time 0, and let $P_0$ be the present value of a dollar to be paid in $T$ years. Discounting this future dollar all the way from $T$ to today using the short rate, we have:

$$P_0 = E\left(e^{-\int_0^T r_t\,dt}\right) \tag{1}$$

  Let me explain this expression in sequence:

  - The reason why we need to do $\int_0^T r_t\,dt$ is because we have to add up all of the future short rates along the path from 0 to $T$. Take the extreme example of a constant short rate $r$. We have $\int_0^T r_t\,dt = r\,T$ and $P_0 = e^{-r\,T}$.
  - We put $\int_0^T r_t\,dt$ onto $e^{-\int_0^T r_t\,dt}$ because the rates are continuously compounded. (You will find that working with $e^x$ and $\ln(x)$ typically gives us a lot of tractability in Finance.)
  - Later on, we will see how $r_t$ is going to be driven by a random risk factor. Because of this, there could be many paths of $r_t$, depending on the random outcomes of the risk factor. And the present value of a future dollar to be paid in year $T$ is an expectation, $E(\cdot)$, taken over all potential random paths of $r_t$ with $t$ running from 0 to $T$.

- **Relating back to Option Pricing:** The calculation in Equation (1) is similar to the calculation of $E^Q\left(e^{-rT}\left(K - S_T\right)\mathbf{1}_{S_T<K}\right)$ in option pricing. The difference is that we do not have to deal with the random variation in $S_T$. But we have to deal with the random variation in the riskfree $r$, which turns out to be more difficult to deal with.

  Instead of fixing a maturity date for this interest rate $r$ (as in yields to maturity), we choose to work with the "short rate" so that this one rate can be used to discount future cashflows over any horizon. We just need to add them up via $\int_0^T r_t\,dt$.

  A by-product of this modeling choice is that we now have to keep track of the entire path of $r_t$ from 0 to $T$ in order to calculate $\int_0^T r_t\,dt$. Remember that when you performed

option pricing via simulation in your Assignment 3, you didn't have to keep track the path of $S_T$ from 0 to $T$. You only needed to know the values of $S_T$. So in order to have one million scenarios of $S_T$, you needed to simulate one million random variables.

To price bonds, however, you need to simulate the entire path of $r_t$ from 0 to $T$. Suppose we decide to discretize the time interval from 0 to $T$ into monthly intervals, then pricing a one-year bond with one million scenarios would involve simulating 12 $\times$ one million random variables; pricing a 10-year bond would involve simulation 120 $\times$ one million random variables. In short, pricing bond is generally more involving than pricing equity options and pricing bond derivatives would be even more challenging. That is why models with closed-form solutions are very useful. Otherwise, we will have to resort to either simulations or solving partial differential equations.

Also notice that to be precise, I should take the expectation in Equation (1) under the risk-neutral measure. For this class, however, let me not make a distinction between the two, just to keep things simple.

- **The Vasicek Model:** In the Vasicek model, the short rate $r_t$ follows

$$dr_t = \kappa\,(\bar{r} - r_t)\;dt + \sigma\,dB_t\,, \tag{2}$$

where, as in the Black-Scholes model, $\sigma\,dB_t$ is the diffusion component with $B$ as a Brownian motion. This model has three parameters:

  - $\bar{r}$: The long-run mean of the interest rate, $\bar{r} = E(r_t)$.
  - $\kappa$: The rate of mean reversion. When $r_t$ is above its long-run mean $\bar{r}$, $\bar{r} - r_t$ is negative, exerting a negative pull on $r_t$ to make it closer to $\bar{r}$. A larger $\kappa$ amplifies this pull of mean reversion and a smaller $\kappa$ dampens it. Conversely, when $r_t$ is below its long-run mean $\bar{r}$, $\bar{r} - r_t$ is positive, exerting a positive pull on $r_t$, again to make it closer to its long-run mean $\bar{r}$.
  - $\sigma$: controls the volatility of the interest rate.

- **Bond Pricing under Vasicek:** Bond pricing under the Vasicek model turns out of be very simple. Let today be time t and let $r_t$ be today's short rate, then the time-$t$ value of a dollar to be paid $T$ years later at time $t + T$ is

$$P_t = e^{A + B\,r_t}\,,$$

where

$$B = \frac{e^{-\kappa T} - 1}{\kappa}$$
$$A = \bar{r}\left(\frac{1 - e^{-\kappa T}}{\kappa} - T\right) + \frac{\sigma^2}{2\kappa^2}\left(\frac{1 - e^{-2\kappa T}}{2\kappa} - 2\frac{1 - e^{-\kappa T}}{\kappa} + T\right)$$

# 2 Calibrating the Model to the Data

- **The Vasicek Model:** As usual, we work with models in order to understand, at a conceptual level, the key drivers in the pricing of a security. Applying the model to the data, we further understand quantitatively how well the model works and what's missing in the model.

  For a one-factor model such as the Vasicek model, we know its limitation even before applying it to the data. In the fixed income market, the level of the interest rates is the number one risk factor in terms of its importance, but it is not the only risk factor.

  In Assignment 4, I ask you to work with a discrete-time version of the Vasicek model by first estimating the model parameters, $\bar{r}$, $\kappa$, and $\sigma$, using the time-series data of 3-month Tbill rates. Basically, I am asking you to calibrate the model only to the time-series information of the short-end of the yield curve, without allowing you to take into account of the information contained in the other parts of the yield curve. Then I ask you to price the entire yield curve. Not surprising, you will find that the calibrated model does not work very well to accommodate the different shapes of the yield curve.

  An alternative approach is to calibrate the model using the yield curve. For example, on any given day, we estimate the model parameters, $\bar{r}$, $\kappa$, and $\sigma$, so that the pricing errors between the model yields and the market yields are minimized. By doing so, the model will do a much better job in matching the market observed yield curve, but it will miss the time-series information. Moreover, you will have one set of parameters per day, which is inconsistent with the assumption that these parameters are constant.

  The better solution is to introduce more factors to the model. For example, instead of forcing the long-run mean $\bar{r}$ to be a constant, we can allow it to vary over time by modeling it as a stochastic process. Instead of forcing the volatility coefficient $\sigma$ to be a constant, we can allow it to vary as another stochastic process. There, you have a three-factor model. The pricing will be more complicated and so will be the estimation. Working with these multi-factor models requires some patience, perseverance, and the

love for the subject matter. Indeed, it is not for everybody.

- **Curve Fitting:** On a topic related to model calibration is yield curve fitting. In this approach, there is no consideration along the time-series dimension. The zero rate $r(\tau)$ of maturity $\tau$ is modeled as a parametric function, which is then used to price all market traded coupon-bearing bonds. On any given day, the parameters in that parametric function will be chosen so that the pricing errors between the model yields and the market yields are minimized. This exercise of yield curve fitting is repeated daily and the model parameters are updated daily as well.

  Figure 2 plots the yield curve during the depth of the 2008 crisis. It uses the Svensson model for curve fitting. The parameters in the Svensson model are first optimized so that the model can price all of the market-traded bonds on December 11, 2008 with minimum pricing errors. Using these parameters, the black line is the corresponding par coupon curve. The blue or purple dots are the market yields for the market-traded bonds. For each dot, there is a companion red "+", which is the model yield for the corresponding bond. In a fast decreasing interest rate environment such as December 2008, most of the existing bonds are premium bonds. As we discussed earlier, with an upward sloping term structure, the yields of these bonds are lower than the corresponding par-coupon yields of the same maturity. That is why most of the red "+"s are below the par coupon curve. If there are many discount bonds being traded at the time, then you will see some red "+"s above the par coupon curve.

  This curve fitting exercise is useful in connecting the yields of different maturities through a parametric function of zero rates. For example, there is quite a bit of overlap in discount rates between a ten-year yield and a ten-year minus one-month yield. The presence of a parametric function of zero rates acknowledges the overlap (ten years minus one month) and the pricing difference between these two yields will be sensitive only to the one-month gap. But the usefulness of a curve fitting exercise stops at this level. If you would like to use a model to help you with derivatives pricing on the yield curve (e.g., Bond options, swaptions, caps/floors, etc), a curve-fitting model will not be helpful at all because it does not take into consideration of how yields vary over time. For derivatives pricing on the yield curve, you need to use dynamic models. The usual approach is to use multi-factor versions of CIR or Vasicek models. Affine models are examples of these multi-factor versions of CIR and Vasicek.

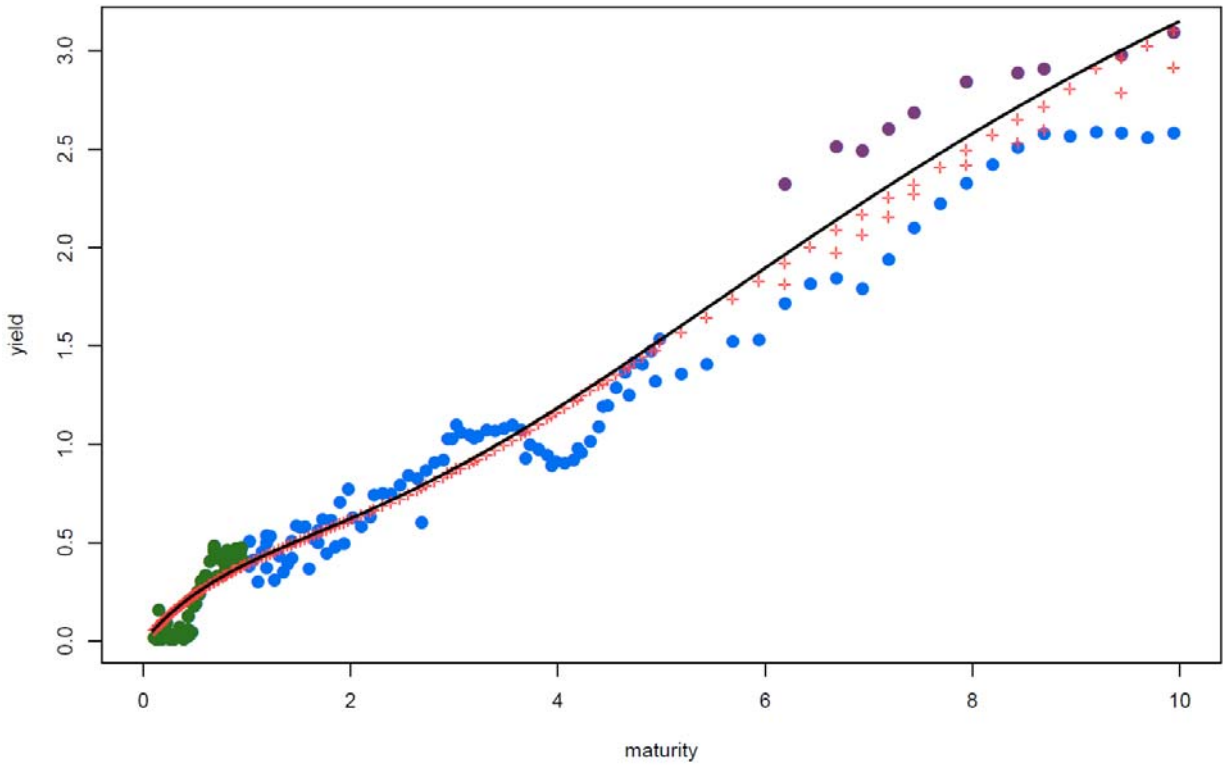


Figure 2: Treasury Yield Curve on December 11, 2008.

# 3 Relative Value Trading with a Term Structure Model

In March 2011, Chifu Huang (a former MIT Sloan Finance professor) came to Prof. Merton's class to give a guest lecture. I found his talk to be very informative and the following is based on one portion of his talk.

- **How to Use a Term-Structure Model to Identify Trading Opportunity:** Relative value trading in the fixed income market does not make a judgment on the level of interest rates or the slope of the curve. It assumes that a few points on the yield curve are always fair. For example, the time-series data on the 10yr, 2yr, and 1-month rates can be used to estimate a three-factor term structure model.

  Recall that in the Vasicek model, the short rate is the only risk factor (i.e., state variable). That is why in your Assignment 4, I ask you to estimate the model using only the 3M Tbill rates. With a three-factor model, we have three risk factors (i.e., state variables) and we need three points on the yield curve to help us estimate the model. Intuitively, the 10yr gives us information about the level of long-term interest rates; the 2yr together with the 10yr informs us about the slope of the curve; and the 1-month Tbill rate captures the short-term interest rate (including expectations on monetary policy in the near term).

  Once you have the model estimated by the time-series data (which is a non-trivial task if you would like to do it properly), this model is going to give you predictions about the level of interest rates across the entire yield curve. You can then compare the model price with the market price to judge for yourself whether or not a market price is cheap or expensive. Once you convince yourself that your model helps you pick up a trading opportunity, you would structure a trade around it. You can buy cheap maturities and sell expensive maturities, and, at the same time, hedge your portfolio so that it is insensitive to the changes of the level or the slope of the yield curve.

  The main judgment call is to understand why your model identifies some maturities as cheap or expensive. If it is due to institutional reasons (which does not show up in your model but does show up in the data), then you can make judgment as to whether or not such institutional reasons will dissipate over time (and how fast).

- **An Example:**

  One example was given by Chifu. In August 1998, Russian defaulted on its local currency debt, and the effect lingered well into September and was later known as the LTCM crisis. As shown in Figure 3, in September 1998, bond markets rallied in

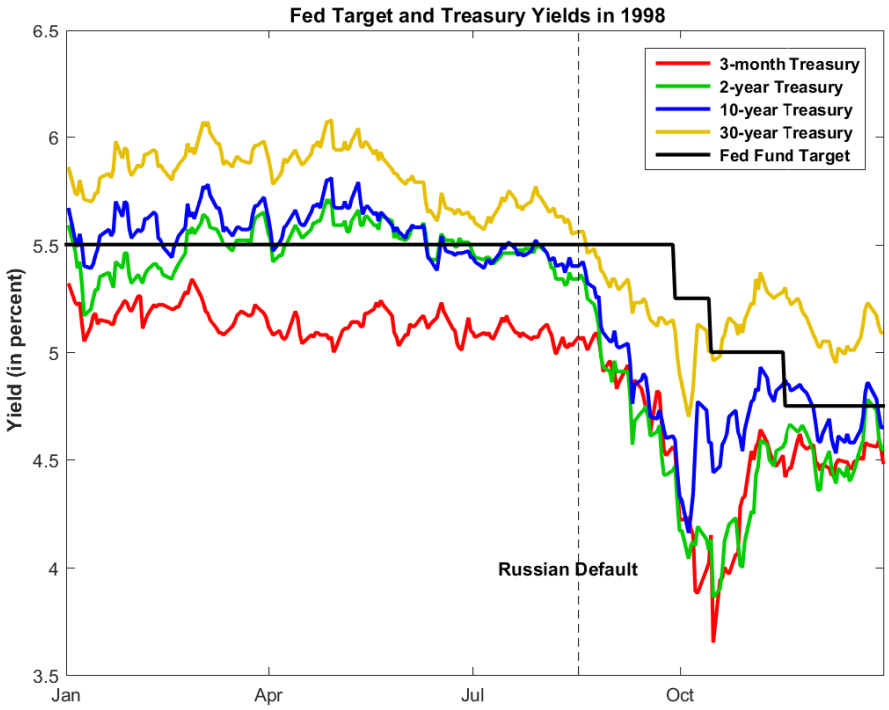


Figure 3: Fed Target and Treasury Yields in 1998.


Figure 4: Cheapness and Richness of US 30-Year Swap Rate Based on a Two-Factor Model.

anticipation of a rate cut. On September 29, the Fed cut the fed funds target rate by 25 bps.

Figure 4 is a slide presented by Chifu in his talk. In September 1998, his two-factor model picks up a trading opportunity regarding the 30yr bond. According to the model, the market price for the 30yr bond is cheap relative to the model price. The deviation between the data and the model was at the range of 10 to 20 bps. The 30-year rate was around 5.5% at that time, implying a modified duration of about 15 years. So a 10 bps price deviation in 30yr would translate to 10 bps $\times$ 15 = 150 bps in bond return. And a 20 bps deviation will translate to 3% in bond return.

So what are the reasons for this cheapening of 30yr? It is because residing over the 30yr region are pension funds and life insurance companies who are either inactive "portfolio rebalancers" or rate-targeted buyers. As a result, the rally that happened in the rest of the yield curve didn't find its way to the 30yr region. There is a lag in how information (regarding an impending rate cut) gets transmitted to this region. As you can see from Figure 3 and 4, it was only after the Fed's rate cut on September 28 when the 30yr yield was brought back in alignment with the rest of the yield curve.

Jun Pan
junpan@mit.edu
MIT Sloan School of Management
15.433-15.4331 Financial Markets, Fall 2015
253-3083
E62-624

## Class 21: Fixed Income, Yield Curve

This Version: November 30, 2016

# 1 Factors Influencing the Yield Curve

- **The Yield Curve:** The Treasury yield curve is the best way to summarize the market prices of Treasury bonds, just like the implied-vol curves in the options market. In options, the vol curves become a three-dimensional surface because of an option could vary in its moneyness as well as time to expiration. In bonds, the yield curve remains a two-dimensional curve: a plot of yield against maturity.

  Of course, bonds of the same maturity also vary in their "moneyness": new bonds are issued at par with $y = c$ and $P = \$100$; old bonds issued during high interest-rate environment are premium bonds with $c > y$ and $P > \$100$; bonds issued during extremely low interest-rate environment will eventually become discount bonds with $c < y$ and $P < \$100$. Because of this, when we talk about yield curve, we need to be more specific. In general, for coupon bonds, we usually use the par curve: the yields for par coupon bonds. For a given maturity, the yield of a par coupon bond will be located ... exactly on the curve, while the yields for discount/premium bonds will be close to the curve but slightly off.

  With an upward sloping yield curve, the yield of a premium bond sits below the par curve while the yield of a discount bond sits above the par curve. This is, because the premium bond, with relatively higher coupon payments $c^{\text{premium}} > c^{\text{par}} > c^{\text{discount}}$, puts a relatively higher weight on the yields of shorter maturities. With an upward sloping yield curve, this translates to a slightly lower yield. Overall, however, the differences are not huge. I would encourage you to go through the math yourself to verify this intuition and gauge the magnitude.

  In doing these calculations, there is always a curve in the background that guides our intuition. That is the zero curve, which is effectively the collection of discount functions over different maturities. Having this zero curve is useful in discounting cashflows and

we can price bonds of all maturities and varying coupon rates, using the discount function dictated by the zero curve. As a result, this zero curve enforces the pricing of bonds of all maturities to be internally consistency. Now the question is where do we get a zero curve in the first place? Most of the bonds in the market are coupon-bearing bonds, we do not observe the zero curve directly. So the common practice is to build a zero curve from market prices of coupon-bearing bonds. In fact, this task of yield curve fitting should be a very basic skill for a fixed-income person. If I have time for the next class (on term structure modeling), I'll talk more about the exact approach. By the way, using the intuitive developed earlier about premium/discount bonds, we know that, with an upward sloping yield curve, the zero curve sits above the par curve.

- **Factors Influencing the Yield Curve:** Our discussion so far focuses on the internal consistency of bond pricing. We imagine that there is a zero curve and ask the pricing of all coupon-bearing bonds to be consistent with this zero curve. The first question you would ask is: what are the factors influencing this zero curve? In an environment of constant interest rate, this zero-curve would always be flat. Then there is not too much to talk about. In practice, the curve is not flat and interest rates are not constant. So what can we learn about them? We will try to answer this question in today's class.

  Once we are happy with the answers to the first question, we will ask the second question, which is also very interesting. With a zero curve (or even a sophisticated term structure model), we price all coupon-bearing bonds traded in the market. How good is our curve (or model) in pricing all of these bonds? What are we to learn when some bonds are mis-priced by a curve (or model)? We will try to answer this question in our next class.

  As you can see, both questions focus on the same issue: what are the factors influencing bond pricing in the market place? I split the question into two so that we can answer this important question in two steps. First, we address the economic factors influencing the yield curve. In a way, these factors are more macro and systematic, affecting every "body" on the yield curve. Second, we address the institutional reasons affecting the yield curve. These factors are more localized and idiosyncratic.

- **Movements and Co-Movements in Yield Curve:** Figure 1 plots the time-series of Treasury yields of the representative maturities: 3M, 2Y, 5Y, 10Y, and 30Y. As you can see, most of the time, the yield curve is upward sloping. In a few occasion, the yield curve becomes flat or even inverts to a downward sloping curve.

  It is also evident from the plot that there is quite a bit of comovement in yields across

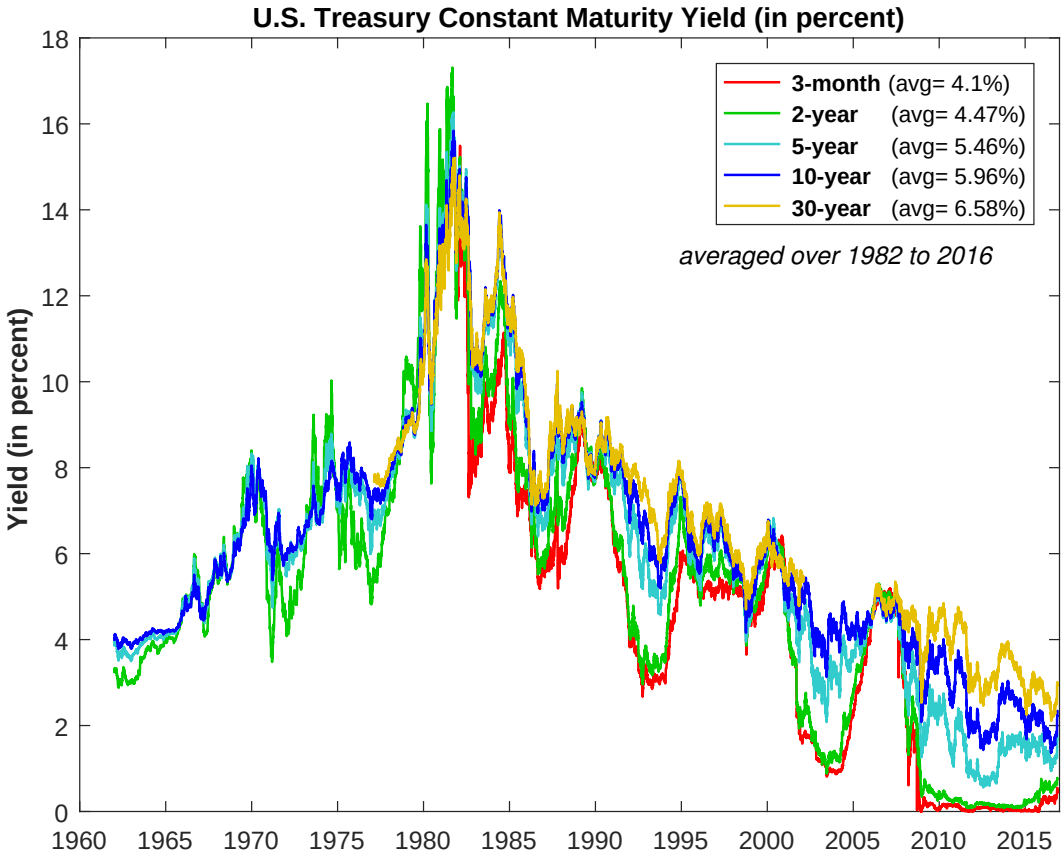


Figure 1: Time-Series of Treasury Constant Maturity Yields.

Table 1: Comovement in Yields, Daily Data from 1982 to 2015

| corr in yields (%) | | | | | |
|---|---|---|---|---|---|
| | 3M | 2Y | 5Y | 10Y | 30Y |
| 3M | 100.0 | 98.57 | 96.19 | 93.61 | 90.90 |
| 2Y | **98.57** | 100.0 | 99.18 | 97.54 | 95.47 |
| 5Y | 96.19 | **99.18** | 100.0 | 99.46 | 98.19 |
| 10Y | 93.61 | 97.54 | **99.46** | 100.0 | 99.57 |
| 30Y | 90.90 | 95.47 | 98.19 | **99.57** | 100.0 |
| corr in daily changes in yields (%) | | | | | |
| | 3M | 2Y | 5Y | 10Y | 30Y |
| 3M | 100.0 | 57.31 | 46.87 | 40.18 | 35.15 |
| 2Y | **57.31** | 100.0 | 90.29 | 82.17 | 72.90 |
| 5Y | 46.87 | **90.29** | 100.0 | 94.07 | 85.74 |
| 10Y | 40.18 | 82.17 | **94.07** | 100.0 | 93.71 |
| 30Y | 35.15 | 72.90 | 85.74 | **93.71** | 100.0 |

different maturities. This is better summarized in Table 1. The pairwise correlations between yields of different maturities are well above 90%. Given the visible time trend and the persistence in yield (the auto-correlation in yield is close to 1), it is more meaningful to measure the comovement in changes in yields. After all, it is the surprise components (i.e., the random shocks) in yield that interest us the most. Measuring the pairwise correlations between daily changes in yields, we still find substantial co-movements. Within the Treasury bonds and notes, the correlations between the two nearest maturities are above 90%. The comovement becomes relatively weaker as the maturities are further apart. But even for the 2Y and 30Y bonds, the correlation is around 70%. The connection between the Treasury bills and the rest of the yield curve is relatively weaker but still substantial: the correlation between 3M TBill and 2Y bond is about 57%.

Overall, we can see that the Treasury yield curve is an inter-connected curve. It is not a curve with its individual components moving around freely without any regard for other parts of the curve. In this sense, the curve is a tight family of individual members. But it is also not a curve with its individual components moving in exactly the same pace. There is some internal consistency and relationship. The closer the maturity, the stronger the relationship. Let's try to figure out the economic factors that drive these movements and comovements.

- **Monetary Policy and Fed Funds Rate:** By far, the most important factor in-

fluencing the yield curve is monetary policy. In the US, monetary policy is carried out by the Federal Reserve through the Federal Open Market Committee (FOMC). In 1977, Congress set explicit objectives for monetary policy: "maximum employment" and "price stability." These two objectives in the Fed's so-called dual mandate are not always in alignment and the committee members of FOMC face the task of making the right decision when these two objectives are in conflict with each other.

The fed funds rate is the main policy tool of the Fed. It is the rate at which depository institutions lend excess reserve balances to each other overnight. Bank reserves are funds that banks hold at the Fed, much like the checking accounts that individuals have at banks. A bank can use its reserve account at the Fed for making or receiving payments from other banks, as well as a place to hold extra cash. Banks are legally required to hold a minimum level of reserves. If a bank finds itself with reserve balances in excess of the required minimum level, it often lends the excess reserves out to other banks in the so-called fed funds market, in the form of an unsecured overnight loan. And the interest rate of this private loan is the fed funds rate.

Quoting the former Chairman Ben Bernanke, "Although the federal funds rate is a private rate between banks, the Fed was able to control it indirectly by affecting the supply of funds available to banks. More precisely, the Fed managed the funds rate by affecting the quantity of bank reserves."

"The Fed was able to affect the quantity of bank reserves in the system, and thereby the federal funds rate, by buying or selling securities. When the Fed sells securities, for example, it gets paid by deducting their price from the reserve account of the purchaser's bank. The Fed's securities sales consequently drain reserves from the banking system. With fewer reserves available, banks are more eager to borrow from other banks, which puts upward pressure on the federal funds rate, the interest rate that banks pay on those borrowings. Similarly, to push down the federal funds rate, the Fed would buy securities, thereby adding to reserves in the banking system and reducing the need of banks to borrow from each other."

Effectively, the Fed's balance sheet is like a gigantic balloon attached to the entire US banking system. If the Fed feels that the economy is at the risk of overheating (i.e., the risk of high inflation), it will suck some air out of the system by selling securities into the system and therefore draining cash out of the banking system. If the Fed feels that the economy is performing poorly (i.e., the risk of high unemployment), it will blow some air into the system by buying securities from the system and therefore replenish the banking system with more cash.

In each of the FOMC meetings, the committee members weight the option of tightening (rate hike), loosening (rate cut), or no action. Over the history of FOMC meetings, the committee members are not always in agreement in terms of the right policy action. Hence the term *hawk*, who puts a higher weight on keeping the inflation low and often biases toward a tighter monetary policy; and *dove*, whose concern with respect to inflation is not as strong and often biases toward a loosening monetary policy. Of course, each policy decision is an "organic" process, with committee members taking into account of the information available to them at the time. If you read the memoirs of the former chairmen (e.g., Greenspan and Bernanke), you will notice that each decision weights heavily in their memory and on their conscience. Such men and women perform a great service to society.

- **Monetary Policy, Historical Experiences:** Figure 2 plots the time-series of Treasury yield curve (2-year and 10-year) along with the historical information that is relevant for our understanding of the monetary policy: inflation rate, GDP growth, and fed fund rates (the black solid line starting in the 1990s).

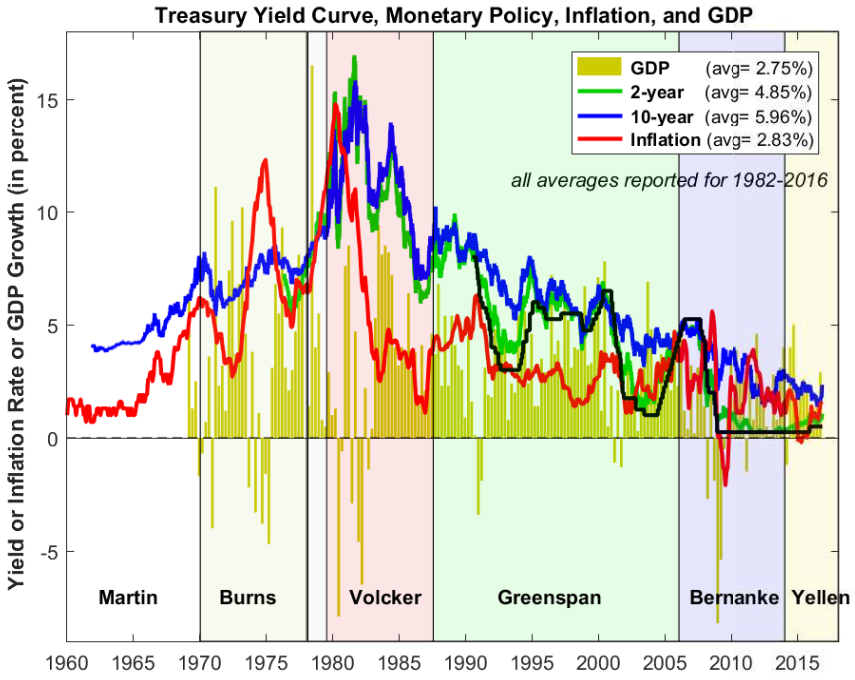


Figure 2: Treasury Yield Curve, Monetary Policy, Inflation Rate and GDP Growth.

Instead of writing about the historical experiences accompanying the events in Figure 2,

let me use the following extended excerpt from Ben Bernanke's book. I find it to be very useful in my own understanding of the US monetary policy, especially for the pre-Greenspan era of Volcker and Burns. It was a period over which I know very little about because ... I was still in China and thought of interest rate as little rectangular stamps collected in a little booklet. Even that remotely related activity was performed only once or twice when my dad dragged a reluctant me, less than ten years of age, to a bank in an effort to educate me on the virtue of being frugal and the benefit of saving.

> Throughout most of the 1990s the Fed presided over an economy with employment growing strongly and inflation slowly declining to low levels. The Fed was thus meeting both parts of its congressional dual mandate to pursue maximum employment and price stability. In contrast, when I arrived at the Fed (August 2002), we saw risks to both sides of our mandate. On the employment side, we had the jobless recovery to contend with. On the price stability side, we faced a problem unseen in the United States since the Depression – the possibility that inflation would fall too low or even tip into deflation, a broad decline in wages and prices.
>
> In the past, the end of a recession had typically been followed by an improving jobs market. But during the two years after the recession that ended in November 2001, the U.S. economy actually lost 700,000 jobs, and unemployment edged up from 5.5 percent to 5.8 percent even as output grew. Many economists and pundits asked whether globalization and automation had somehow permanently damaged the U.S. economy's ability to create jobs. At the same time, inflation had been low and, with the economy sputtering, Fed economists warned that it could fall to 1/2 percent or below in 2003. Actual deflation could not be ruled out.
>
> Worrying about possible deflation was a new experience for FOMC participants. Ever since the end of the Depression, the main risk to price stability had always been excessive inflation. Inflation spiraled up during the 1970s. Paul Volckers Fed ended it, but at a steep cost. Within a few months of Volckers becoming chairman in 1979, the Fed dramatically tightened monetary policy, and interest rates soared. By late 1981, the federal funds rate hit 20 percent and the interest rate on thirty-year fixed-rate mortgages topped 18 percent. As a consequence, housing, autos, and other credit-dependent industries screeched to a halt. A brief recession in 1980 was followed by a

deep downturn in 1981-82. Unemployment crested above 10 percent, a rate last seen in the late 1930s.

After succeeding Volcker in 1987, Alan Greenspan continued the fight against inflation, although he was able to do so much more gradually and with fewer nasty side effects. By the late 1990s, the battle against high inflation appeared to be over. Inflation had fallen to about 2 percent per year, which seemed consistent with Greenspans informal definition of price stability: an inflation rate low enough that households and businesses did not take it into account when making economic decisions.

The Great Inflation of the 1970s had left a powerful impression on the minds of monetary policymakers. Michael Moskow, the president of the Federal Reserve Bank of Chicago when I joined the FOMC (August 2002), had served as an economist on the body that administered the infamous – and abjectly unsuccessful – Nixon wage-price controls, which had attempted to outlaw price increases. (Predictably, many suppliers managed to evade the controls, and, where they couldn't, some goods simply became unavailable when suppliers couldn't earn a profit selling at the mandated prices.) Don Kohn had been a Board staff economist in the 1970s under Fed chairman Arthur Burns, on whose watch inflation had surged. Greenspan himself had served as the chairman of President Ford's Council of Economic Advisers and no doubt shuddered to remember the Ford administrations ineffectual Whip Inflation Now campaign, which encouraged people to wear buttons signifying their commitment to taming the rising cost of living. With Fed policymakers conditioned to worry about too-high inflation, it was disorienting to consider that inflation might be too low. But it was a possibility that we would soon have to take seriously.

- **Fed Funds Rate and Yield Curve:** Beginning in 1994, the FOMC began announcing changes in its policy stance, and in 1995 it began to explicitly state its target level for the fed funds rate. As you can see in Figure 3, this aspect of monetary policy has an immediate impact on the Treasury yield curve, especially on the short end. In many instances, the bond market, in anticipation of the impending rate change, would price the event in advance. For example, on September 13, 2001, when the bond market re-opened on a limited basis after 9/11, the 3M Tbill rate dropped 52 bps from 3.26% to 2.74%, the one-year rate dropped 50 bps from 3.31% to 2.81%, and the two-year rate dropped 54 bps from 3.52% to 2.99%. It was not until four days later, on September

17, the Fed cut the fed fund target rate from 3.50% to 3%.

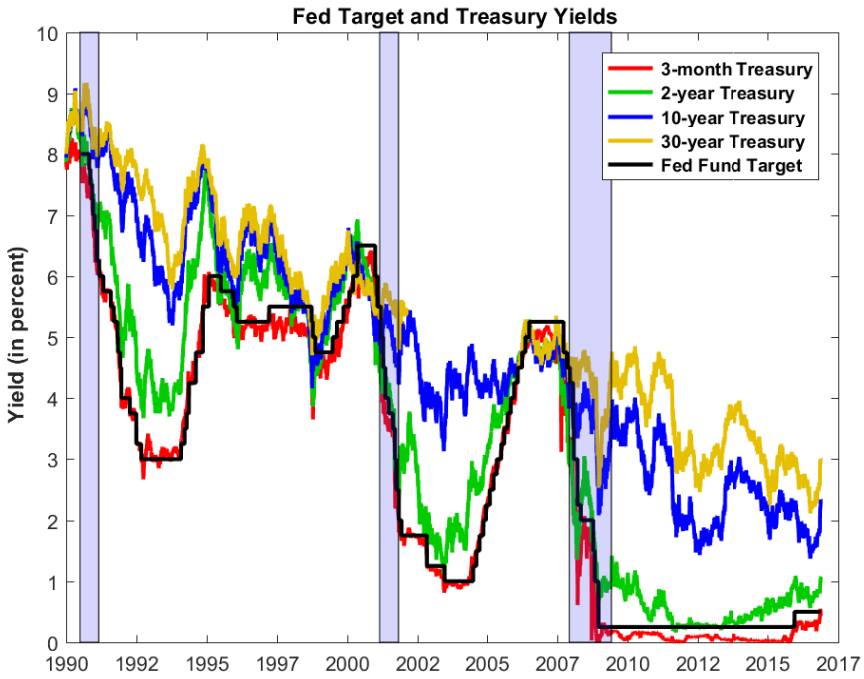


Figure 3: Treasury Yield Curve and Fed Fund Target Rate.

There is also a visible impact on the long-end of the yield curve, although the reactions of the longer end of the yield curve are not one-for-one in magnitude. This of course, makes sense given the impermanent nature of a monetary tightening or loosening. For example, On September 13, 2001, the five-year yield decreased by 38 bps from 4.41% to 4.03% and the ten-year yield decreased by 20 bps from 4.84% to 4.64%. Interestingly, the 30-year yield dropped by only 4 bps from 5.43% to 5.39%. (More on this topic on the 30-year yield next class.)

There were also times when the longer-term interest rates failed to rise after the Fed tightened monetary policy. This happened in 2004-05, the famous Greenspan's "conundrum." Quoting Bernanke again, "In speeches, I tied the conundrum to what I called the 'global savings glut' – more savings were available globally than there were good investments for those savings, and much of the excess foreign savings were flowing to the United States. Additional capital inflows resulted from efforts by (mostly) emerging-market countries like China to promote exports and reduce imports by keeping their currencies undervalued. To keep the value of its currency artificially low

relative to the dollar, a country must stand ready to buy dollar-denominated assets, and China had purchased hundreds of billions of dollars' worth of U.S. debt, including mortgage-backed securities."

Before leaving Figure 3, let me point out one well-known pattern: Prior to each of the three NBER-dated recessions, the yield curve was either very flat or even inverted. It turns out that the slope of the interest rate is a good predictor for future GDP growth.

- **Factors Influencing Monetary Policy:** Also plotted in Figure 3 are the NBER-dated recession periods. The cyclical nature of the monetary policy is obvious in this plot: tightening during economic expansions and loosening during recessions. It is also worthwhile to note that, in order to cause minimal disruption to the markets, the monetary policy applies itself to the market gradually. A typical rate cut/hike is in increments of 25 bps. There were four rate hikes that were 50 bps (twice in 1994, once in 1995 and 2000) and one rate hike of 75 bps (November 1994). There were sixteen rate cuts of 50 bps (three times in 1991-92, nine times in 2001-02 and four times in 2007-08) and three rate cuts of 75 bps (all happened in 2008).

  Given the dual mandate of price stability and maximum employment, it is not surprising that expectations of the rate of inflation, GDP growth, and employment numbers (e.g., nonfarm payroll employment) influence the decision of the policy decision of the FOMC. The Stanford economist John B. Taylor wrote a paper in 1993, linking the policy rate explicitly to inflation rate and GDP. This became the famous Taylor rule and there are various extensions of this rule. Again, if you read the memoirs of Greenspan and Bernanke, you would see that each policy decision is an "organic" process, with committee members taking into account of the information available to them at the time. Having a mechanic rule is useful as a baseline, but cannot be the ultimate answer.

  If you pay attention to the famous Wall Street activity called the "Fed Watch," you will notice market participants use all kinds of signal trying to predict the next policy move. Some macro investors also perform directional trades to express their views and they typically do so using the two-year notes. As such, the two-year yield are considered to be highly sensitive to changes in the Fed's policy outlook. Consequently, the shape of the yield curve (relative to the two-year yield) might contain information about the impending policy move. As you can see in Figure 3, the two-year yield has been increasing quite steadily since the beginning of 2015 in anticipation of the monetary tightening in the end of December 2015.

  In addition, investors also use fed funds futures traded on CME to express their views.

Consequently, the pricing information in this market has been used to extract expectations of future Fed actions. This is a number watched closely by fixed-income traders and global macro investors. Even the Fed tracks this number to gauge the market expectation of their action. According to this calculation the implied probability of a rate hike from the current 25-50 bps to 50-75 bps is about 93.5%.

The market participants are involved in "Fed Watch" because uncertainties in the target rate have a big impact on the markets, not only the bond market but also the stock market (and the currency market). The Fed under chairman Bernanke and chairwoman Yellen has been working hard on Fed transparency in order to better communicate with the market participants in terms of the Fed policy.

- **Quantitative Easing and Operation Twist:** Earlier, we talked about how the Fed can use this gigantic balloon to suck/blow air into the entire banking system by selling/buying securities. Up to the 2008 crisis, the Fed performed monetary policy through affecting the fed funds rate. Starting from late 2008, the Fed employed a policy tool that is highly unorthodox and controversial: purchasing hundreds of billions of dollars of securities directly from the market with the intention to keep the long-term interest rates low.

  After the FOMC meeting on October 29, 2008, the fed funds target rate was at 1% and the 3M Tbill rate was at 62 bps (the Treasury bill rates are lower than the overnight fed funds rate because of the potential counterparty risk involved in the unsecured fed funds loans). When the short-term interest rate reaches close to zero, what to do to bring down the longer-term interest rates in an effort to keep the economic recovery going? One way is to try to convince the market participants that the short-term interest rate will be kept low for a long time. In addition, the Fed also started to to purchase securities in an effort to directly influence the long-term interest rate. On November 25, 2008, the Fed announced plans to perform large scale asset purchases, often referred to as "Quantitative Easing" or QE. As shown in Figure 4, the actual purchases happened in December 2008 for agency bonds (Fannie and Freddie debt) and January 2009 for mortgage-backed securities backed by Fannie Mae, Freddie Mac, and Ginnie Mae. This was later known as QE1, because of it was followed by QE2 and QE3.

  From Figure 4, you can see that the Fed also purchased around $300 billion in Treasury securities during QE1, partly to supplement the reduction in MBS holdings when the mortgages underlying the MBS were paid off (either because of home sales or

refinancings due the decreasing interest rate). We will visit this issue of negative convexity of MBS in a later class.

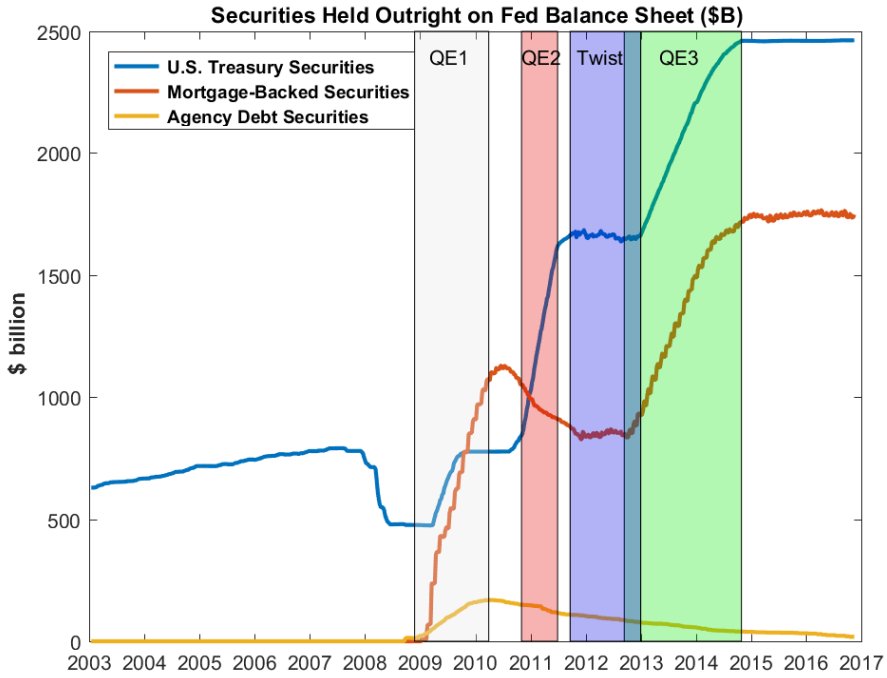


Figure 4: The Fed's Balance Sheet.

QE1 was followed by QE2 and QE3 and a program called "operation twist" in between. The securities purchased through QE2 and QE3 can be seen through the Fed's balance sheet in Figure 4. By now (November 11, 2015), the total market value of securities held outright on Fed's balance sheet is $4.24 trillion, with $2.46 trillion in US Treasury securities. To put these numbers in perspective, let's take a look at some other numbers. According to this Treasury website, as of August 2015, foreign holdings of the Treasury securities totals to $6.099 trillion with China holding $1.27 trillion and Japan holding $1.197 trillion. According to the World Bank, the 2014 GDP is $17.419 trillion for the US, $10.360 trillion for China, $4.601 trillion for Japan, $3.852 trillion for Germany, and $2.942 trillion for the UK.

Figure 5 plots the Fed's holdings of Treasury securities by maturity. As shown Figure 5, during "Operation Twist," the overall Treasury holding by the Fed remains nearly constant in market value. But the maturity of Fed's holdings went through a big change. As shown in Figure 5, the Fed was actively selling Treasuries securities

maturing in 1-5 years and buying longer maturity bonds (5-10 years and longer than 10 years). Effectively, the Fed was increasing the duration of its Treasury portfolio without having to expand its balance sheet, in an effort to influence the long maturity yields so as to reduce the cost of credit for mortgage loans and corporate bonds.

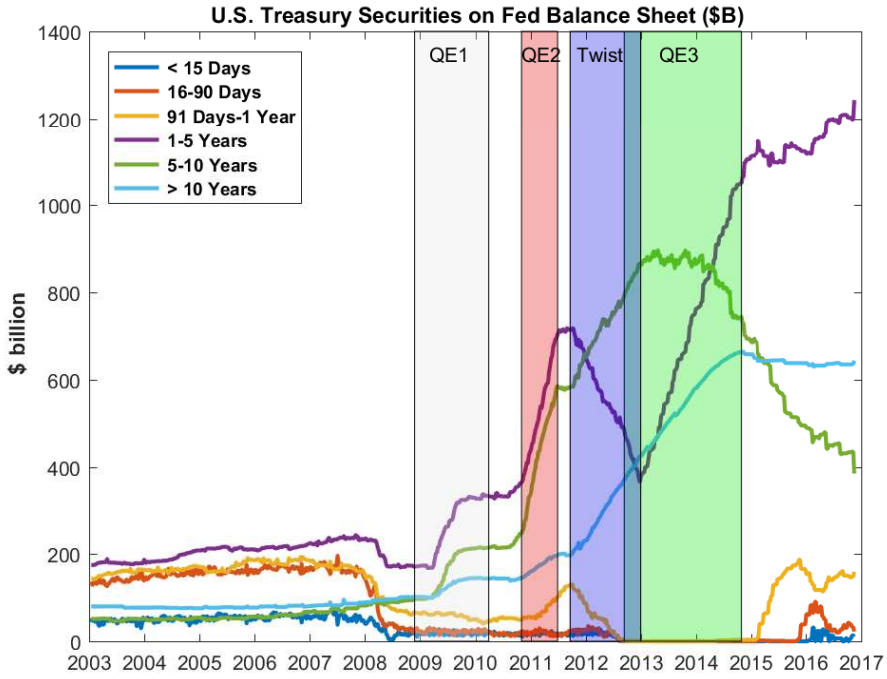


Figure 5: US Treasury Securities on Fed Balance Sheet, Maturity Composition.

The unconventional QE programs and the burgeoning Fed balance sheet were certainly not without risk. To put the policy thinking in perspective, let's take a look at the macro variables from 2008 to 2015 (see Figure 6). Prior to QE2, around October 2010, the unemployment rate was at 9.6%. The last time the unemployment rate was this high was during 1982-83 after the monetary tightening by Chairman Volcker. By contrast, the inflation was low at 1.1% in October 2010. Prior QE3, around August 2012, the employment rate was at 8.1% and the inflation was at 2%. It was clear that at the time, the Fed felt that the unemployment rates were too high (and inflation was not an issue of big concerns) and the economy needed help ... from somewhere. And the Fed's decision at the time was to step up and provide that help.

The economy has certainly been doing relatively better since then. As shown in Figure 6, by the end of QE3, the unemployment rate has been decreasing steadily to 5.7%

and the GDP growth was at 4.3%. Right now (October 2015), the unemployment rate is at 5% and the GDP growth has been uneven: 1.5% for the third quarter and 3.9% for the second quarter. Overall, however, it is difficult to quantify the effect of the QE programs. How do you evaluate the counterfactual of an economy without QE programs? This, of course, is what differentiates Economics from Physics, where you can do controlled and repeated experiments.

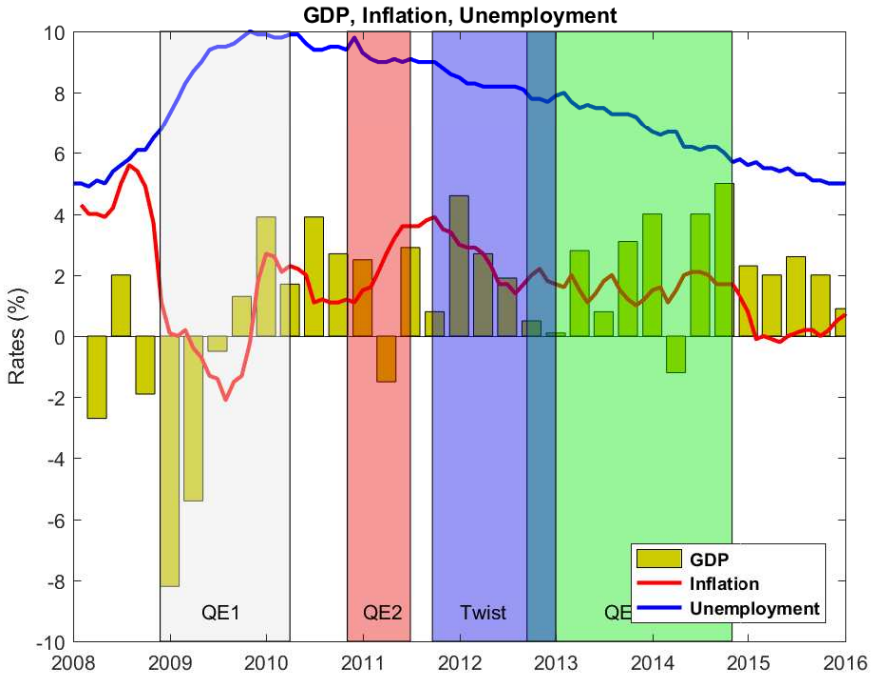


Figure 6: GDP, infation, and unemployment rate since 2008.

At the time, two major concerns about the QE programs were hyperinflation and sharp dollar depreciation. As shown in Figure 6, the inflation rate has in fact been unusually low in recent years with the 2015 numbers hovering around 0. If you follow the currency market, you would know that in recent years the dollar has been strengthening against most currencies, and this is true even before the recent election). So what were the reasons? Let me quote Ben Bernanke again:

> That idea (hyperinflation and sharp dollar depreciation) was linked to a perception that the Fed paid for securities by printing wheelbarrows of money. But contrary to what is sometimes said (and I said it once or twice myself, unfortunately, in oversimplified explanations), our policies did not in-

> volve printing money – neither literally, when referring to cash, nor even metaphorically, when referring to other forms of money such as checking accounts. The amount of currency in circulation is determined by how much cash people want to hold (the demand goes up around Christmas shopping time, for example) and is not affected by the Fed's securities purchases. Instead, the Fed pays for securities by creating reserves in the banking system. In a weak economy, like the one we were experiencing, those reserves simply lie fallow and they don't serve as 'money' in the common sense of the word.
>
> As the economy strengthened, banks would begin to loan out their reserves, which would ultimately lead to the expansion of money and credit. Up to a point, that was exactly what we wanted to see. If growth in money and credit became excessive, it would eventually result in inflation, but we could avoid that by unwinding our easy-money policies at the appropriate time. And, as I had explained on many occasions, we had the tools we needed to raise rates and tighten monetary policy when needed. The fears of hyperinflation or a collapse of the dollar were consequently quite exaggerated. Market indicators of inflation expectations – including the fact that the U.S. government was able to borrow long-term at very low interest rates – showed that investors had great confidence in the Fed's ability to keep inflation low. Our concern, if anything, was to get inflation a little higher, which was proving difficult to accomplish.

Finally, Figure 7 looks at the impact of the unconventional QE policy tools on the level as well as the slope of the yield curve. Again, causality is difficult to establish because we need to know the counterfactual of what would have happened if the Fed had not installed these policies. Also, the issue is further complicated by markets' anticipations at the time as well as the endogenuity of the decision itself. All in all, however, these policy actions seem to be effective in keeping the long-term interest rate low.

- **Why So Much on QEs?** If you feel that I am writing too much here on quantitative easings (more than you need to know), I agree. But more information is always better than no information, right? Rest assured, I'll not ask you to present the pro/con of the QE programs in the final exam.

  I am recounting the events of 2010-2012 regarding quantitative easing for two reasons. First, these were really important events in the fixed income market. By going through the Fed's balance sheet, you get a better sense as to how the Fed's open market

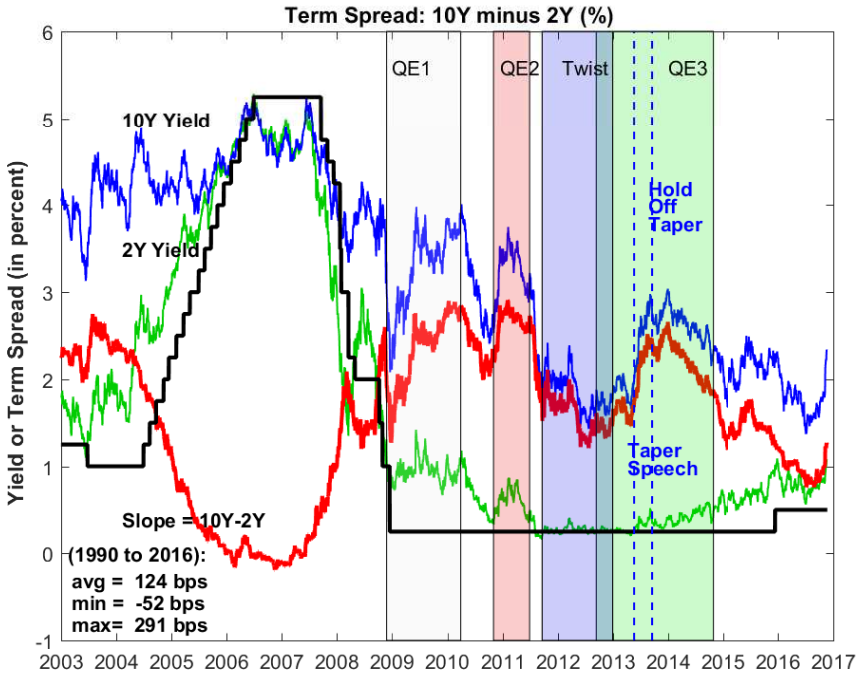


Figure 7: Treasury Yield Curve and Fed Fund Target Rate.

operation actually works. At least that was really helpful for me. The textbook information can be dry sometimes. Having plots like those in Figure 4 and 5 adds texture to my understanding. Second, I lived through that period in 2010-12 listening to many criticisms and derision against the QE programs. I am not a macro-economist and have not been trained in that field. And my thinking at the time was confused by many voices competing for attention. Personally, I find it is useful to read through the above two paragraphs written by Bernanke and look at the numbers for myself. So I thought I would share my readings with you.

I would not be surprised if, for each argument presented by Bernanke in his book, there is a counter argument. Honestly, the writings and thinking of some macro-economists are so complicated that they add more confusion than clarity. In my opinion, truth is always simple. It is the false that needs decoration. Complicated writing comes from a crowded and clouded mind. Unfortunately, in our field, complicated and convoluted thinking is often awarded with a premium because it is an exercise of a high IQ. In any case, for whatever it is worth, I appreciate the clear writing and thinking of chairman Bernanke.

# 2 Statistical Analysis of the Yield Curve

By now, we are comfortable with yield curves and have an intuitive understanding of the various factors influencing the short- and long-end of the yield curve. Let's now move on to quantify these random factors. Not surprisingly, the first risk factor that will show up through our analysis is the risk involved with duration. Second, we also noticed earlier that the entire yield curve does not move in tandem as in synchronized swimming. In particular, the long-end of the yield curve might not move entirely in parallel to the short-end of the yield curve. This points to the fact that the slope of the yield curve is not a constant. In fact, the slope becomes our second random factor. Finally, there might be some freedom in how the middle portion of the yield curve moves in relation to the short- and long-end. This observation gives rise to a third random factor called curvature.

It would not be surprising that market participants have long recognized the importance of these factors influencing the yield curve. But the concept of level, slope, and curvature was formally introduced in the 1991 paper by Litterman and Scheinkman, when both professors were working at Goldman Sachs. They identified these three common factors in the movements of yield curve through principal component analysis (PCA). In assignment 3, you get the chance to do this analysis yourself. The main difference is that their analysis is done in the yield space while your analysis will be done in the return space.

I'll go over this exercise in the yield space here in the notes.

- **Variance-Covariance Matrix:** Table 2 reproduces the content from Table 1 with the addition of 1Y yield. From Figure 1, we also notice that the 30Y yields were absent from February 19, 2002 to February 8, 2006 because the Treasury department suspended new issuance of 30-year bonds. In calculating the variance-covariance matrix, we will have to skip that specific period because of the missing 30-year bonds.

Table 2: Correlation and Standard Deviation of Daily Changes in Yields (1982 to 2015)

| corr (%) | 3M | 1Y | 2Y | 5Y | 10Y | 30Y |
|---|---|---|---|---|---|---|
| 3M | 100.0 | 72.72 | 57.31 | 46.87 | 40.18 | 35.15 |
| 1Y | **72.24** | 100.0 | 87.90 | 78.18 | 70.44 | 63.06 |
| 2Y | 57.31 | **87.90** | 100.0 | 90.29 | 82.17 | 72.90 |
| 5Y | 46.87 | 78.18 | **90.29** | 100.0 | 94.07 | 85.74 |
| 10Y | 40.18 | 70.44 | 82.17 | **94.07** | 100.0 | 93.71 |
| 30Y | 35.15 | 63.06 | 72.90 | 85.74 | **93.71** | 100.0 |
| std (bps) | 8.06 | 6.95 | 6.96 | 7.19 | 6.90 | 6.30 |

Let Cov be the variance-covariance matrix of the daily changes in yields for maturities 3M, 1Y, 2Y, 5Y, 10Y, and 30Y:

$$\text{Cov}(i,j) = \text{Corr}(i,j) \times \sigma_i \times \sigma_j \,,$$

where $\sigma$ is the standard deviation of the daily changes in yield.

- **Eigenvalue Decomposition:** Taking Cov as an input, we perform the eigenvalue decomposition. Let's first go through the calculations and then come back to understand what is really going on. The eigenvalue decomposition will give us two inter-related outputs. First, the eigenvalue E is a vector of six eigenvalues. This is because the dimension of the variance-covariance matrix is 6, one for each maturity. As shown in Table 3, we order the eigenvalues in the order of their magnitude. We call the eigenvalue with the largest magnitude PC1 (principal component one), the second PC2, and so on. The magnitudes of the eigenvalues might not be meaningful for you now, but it will be.

Table 3: Eigenvalues and Eigenvectors

| **Eigenvalues E** | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
|---|---|---|---|---|---|---|
| **E** ($\text{bps}^2$) | 226.99 | 50.14 | 13.77 | 5.45 | 2.86 | 1.47 |
| **E/sum(E)** (%) | 75.49 | 16.68 | 4.58 | 1.81 | 0.95 | 0.49 |

| **Eigenvectors D** | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
|---|---|---|---|---|---|---|
| 3M | 0.3630 | -0.8017 | 0.4347 | 0.1876 | -0.0365 | 0.0006 |
| 1Y | 0.4182 | -0.2371 | -0.4682 | -0.6806 | 0.2939 | 0.0016 |
| 2Y | 0.4351 | 0.0257 | -0.5134 | 0.3309 | -0.6505 | 0.1176 |
| 5Y | 0.4513 | 0.2493 | -0.0709 | 0.4572 | 0.5076 | -0.5124 |
| 10Y | 0.4176 | 0.3430 | 0.2837 | 0.0418 | 0.2271 | 0.7577 |
| 30Y | 0.3550 | 0.3472 | 0.4926 | -0.4258 | -0.4242 | -0.3866 |

Second, associated with each eigenvalue is a vector, called eigenvector. There are six eigenvalues. So there are six eigenvectors, one for each eigenvalue. Putting these six eigenvectors together, we have a matrix D that is 6 by 6, as shown in Table 3. Let's now go over the first three PCs:

– **Level:** As shown in Table 3, associated with PC1 is the first eigenvector:

$$D^{PC1} = \begin{pmatrix} 0.3630 \\ 0.4182 \\ 0.4351 \\ 0.4513 \\ 0.4176 \\ 0.3550 \end{pmatrix},$$

which is a vector of six, one for each maturity. So effectively, the first PC is close to an equal-weighted portfolio of all six yields (or daily changes in yields, to be more precise). This factor corresponds to a movement in the yield curve when all six yields move up and down in tandem or in parallel. In other words, it captures the level movement and the best measure for exposure to this level risk is duration.

– **Slope:** Associated with PC2 is the second eigenvector:

$$D^{PC2} = \begin{pmatrix} -0.8017 \\ -0.2371 \\ 0.0257 \\ 0.2493 \\ 0.3430 \\ 0.3472 \end{pmatrix},$$

which is a long/short portfolio along the maturity dimension. It is long long-term yield and short short-term yield. It really does not matter which end of the yield curve is being long, as long as the weights on the long-end are opposite to the weights on the short-end. Naturally, you think "slope."

– **Curvature:** Associated with PC3 is the third eigenvector:

$$D^{PC3} = \begin{pmatrix} 0.4347 \\ -0.4682 \\ -0.5134 \\ -0.0709 \\ 0.2837 \\ 0.4926 \end{pmatrix},$$

which is again a long/short portfolio along the maturity dimension, but it is long both short- and long-end of the yield curve, and short the middle part of the yield curve. Again, the exact sign of long/short does not really matter as long as the weights on the short- and long-end are opposite to the weights on the middle portion of the yield curve. So this reason, this factor is called "curvature."

Figure 8 summarizes the first three PCs in a plot, which might be more intuitive for us to see the meaning of level, slope, and curvature.

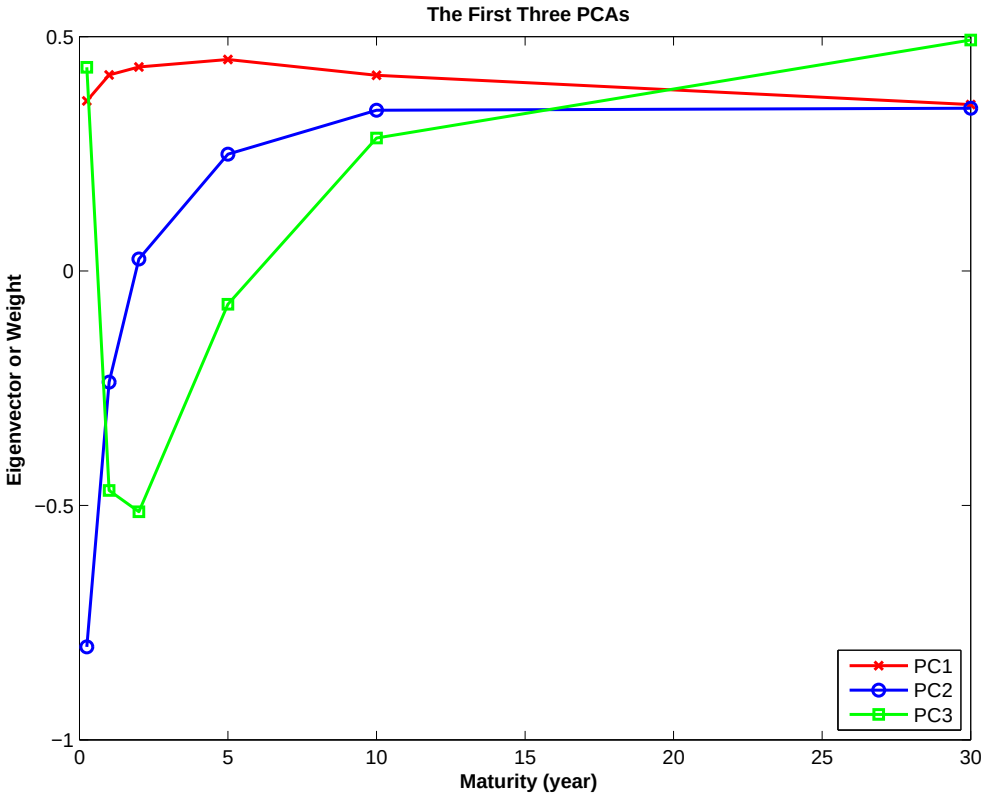


Figure 8: Level, Slope, and Curvature.

- **Relative Importance of the PCs:** We focus on the first three PCs because of their relative importance. To see this, let's go back to the eigenvalues in Table 3. By construction, the eigenvalue associated with PC1 is the highest in magnitude. Let's now construct a time-series of PC1 using the weights subscribed in $D^{PC1}$ (avoiding the 2002-2006 period because of the missing 30-year yields). The standard deviation of this portfolio turns out to be 15.07 bps, and the variance is ... 226.98 (bps$^2$). You can repeat the same exercise for all other PCs. In short, the $n$-th eigenvalue is in fact the variance of the $n$-th PC.

What is cool about the eigenvalue decomposition is that it transforms the original data (with correlated yields) into six independent random factors: PC1, PC2, etc. (Please double check this statement by constructing time-series of PC1 and PC2 and calculate their correlation.) As a result, working with mutually independent PCs is more convenient than working with correlated yields. Because all six factors are independent, we can add now all six eigenvalues into sum(E) and use it as a normalizing factor for E. As shown in Table 3, the first PC accounts for 75.49% of the total variance, the second PC accounts for 16.68%, and the third PC accounts for 4.58%. Adding all three together, we see that they account for 96.75% of the total variance. This is why most of the term structure models use three factors. This is also why duration hedging, which is a hedge against PC1, is the most important form of hedging in the fixed income market.

Once a portfolio is hedged with zero duration, then the slope exposure becomes the most important risk. Once a portfolio is hedged with duration and slope, then you worry about curvature exposure. In the old days, there are butterfly trades which are duration and slope neutral, and are structured so that the main exposure is the curvature risk. Of course, you need to be a fixed-income nerd to get this deep into the yield curve trades.

- **More on the Eigenvectors D:** By now, we understand that there are six eigenvectors and putting them together gives us a $6 \times 6$ matrix. Each eigenvector is a vector of portfolio weights (not normalized) on the six maturities.

  Let's now take a closer look. Let's start with the important observation that all six PCs are independent. So pick any two PCs, say PC1 and PC2, and their correlation will be zero. As mentioned earlier, this is why eigenvalue decomposition is useful. It gives us independent factors. Let's use the matrix notation for the following calculations. First we know that

$$PC1_t = \left(D^{PC1}\right)^{\intercal} \times \Delta\, y_t\,,$$

  where $\Delta y_t$ is the vector of daily changes in yields for the six maturities and $\left(D^{PC1}\right)^{\intercal}$ is the transpose of $D^{PC1}$. Of course, we also know that

$$PC2_t = \left(D^{PC2}\right)^{\intercal} \times \Delta\, y_t\,.$$

  So if $\mathrm{cov}\left(PC1_t, PC2_t\right) = 0$, then it must be that

$$\left(D^{PC1}\right)^{\intercal} \times D^{PC2} = 0\,.$$

Applying this logic pairwise to all maturities, you will be convinced that

$$D^{\intercal}\, D = I\,,$$

where $I$ is an identity matrix of dimension $6 \times 6$, with diagonal terms equaling 1 and off-diagonal terms equaling zero. In other words,

$$D^{-1} = D^{\intercal}\,.$$

If you don't believe me, just try it out using Excel or Matlab.

- **Running Regressions:** Now let's take a look at Table 4, where I report the following regression results:

$$\Delta y_t = a + \beta^{PC1}\, PC1_t + \beta^{PC2}\, PC2_t + \beta^{PC3}\, PC3_t + \epsilon_t\,.$$

Knowing that all three PCs are independent, we can calculate the individual R-squared for each PC and add them together to get the total R-squared of the regression.

Table 4: Regressing $\Delta y$ on the First Three PCs

| maturity | PC1 $\beta$ | PC2 $\beta$ | PC3 $\beta$ | PC1 R2 (%) | PC2 R2 (%) | PC3 R2 (%) | Total R2 (%) |
|---|---|---|---|---|---|---|---|
| 3M | 0.3630 | -0.8017 | 0.4347 | 46.06 | 49.63 | 4.01 | 99.70 |
| 1Y | 0.4182 | -0.2371 | -0.4682 | 82.18 | 5.83 | 6.25 | 94.26 |
| 2Y | 0.4351 | 0.0257 | -0.5134 | 88.67 | 0.07 | 7.49 | 96.23 |
| 5Y | 0.4513 | 0.2493 | -0.0709 | 89.46 | 6.03 | 0.13 | 95.62 |
| 10Y | 0.4176 | 0.3430 | 0.2837 | 83.17 | 12.39 | 2.33 | 97.89 |
| 30Y | 0.3550 | 0.3472 | 0.4926 | 72.04 | 15.22 | 8.41 | 95.66 |

First, you can see that PC1 remains the most important random factor, explaining the daily changes in yields with very high R-squared's. For the two extreme ends of the yield curve (3M and 30Y), the explanatory power is relatively weaker. This is where PC2 picks up. In particular, PC2 contributes quite a bit in explaining the movements in the short-end of the yield curve. Adding all three PC factors, we can explain the random variations in daily changes in yields with R-squared's that are well above 90%.

Second, take a look at the regression coefficients $\beta$'s. What do you see? Compare $D^{PC1}$ with the beta coefficients on PC1, there are identical! Likewise for $D^{PC2}$ and

$\beta^{PC2}$, and $D^{PC3}$ and $\beta^{PC3}$. Can you prove this result mathematically? (No worries, I'll not ask you to do this proof in the exam.)

- **More on PCA:** What we've talked about so far in this section is statistical based. The yield curve is well suited for a statistical analysis like PCA. Once you understand the mechanics of the PCA, it will be instructive for you to go back to the economic drivers for these common risk factors in the fixed-income market.

  More broadly, the PCA approach can also be used in many markets where the observables are correlated due to some common factors. For example, applying PCA to international equity returns, the first PC will be a world index with roughly equal weight on all countries. The second PC will be a long/short portfolio across the two most representative regions (which could change over time).

  Whatever you might do with PCA, just be reminded that this is simply a statistical tool that helps you extract mutually independent factors, and the importance of the factors are ordered by their variances (i.e., eigenvalues). Also remember that the only input for the eigenvalue decomposition is the variance-covariance matrix. Use this tool effectively for the your desired objective. All all is done, take the extra step to understand the economic and institutional drivers for the extracted factors.

Appdendix

# 3 Yield Curve Fitting

The Treasury bond market is a highly liquid market. Everyday, we observe transaction prices of bonds of different coupon rates, and with different remaining maturities. Usually, this information is fed into a yield curve program and the bond prices are transformed into yield curves. There are three yield curves that we should pay attention to: the forward curve, the zero-coupon curve (also called spot curve), and the par-coupon curve.

A good starting point of a yield curve program is the Nelson and Siegel (1987) model.

Jun Pan
junpan@mit.edu
MIT Sloan School of Management
15.433-15.4331 Financial Markets, Fall 2015
253-3083
E62-624

# Class 23: Fixed Income, Interest Rate Swaps

This Version: December 1, 2016

## 1 Interest Rate Swaps

- **Fixed and Floating:** A USD interest rate swap is a private agreement between two counterparties to exchange cashflows in US dollar. One counterparty receives fixed and pays floating, while the other counterparty pays fixed and receives floating. For convenience, let's call one counterparty receiver and the other payer, with reference, in both cases, to their activities on the fixed leg. Later, we will see that the receiver is long duration and the payer is short duration.

  Figure 1 is a Bloomberg screenshot made for me by a former MBA student. It provides a nice description in terms of the fixed and floating legs of an interest rate swap.


Figure 1: Description of a Standard Swap Quote from Bloomberg.

**The fixed leg** is referenced to the swap rate. For example, on November 19, 2015,

the five-year swap is quoted at 1.55%, implying a fixed payment of 1.55%/2 × notional amount, to be delivered every six month, from the payer to the receiver, until the end of contract maturity, which, in this case, is year five. This structure is exactly the same as the coupon payments in Treasury bonds, and this is so by design. The notional amount in the swap market is typically quite large, say $10M per contract. For our analysis, let's make the notional amount $100 so as to match the $100 principal amount we've been using for Treasury bonds. By now, it should be obvious that the notional amount or the principal amount is scalable and its absolute value does not really matter for our analysis. Of course, it will matter when we think in terms of the size of our portfolio and calculate profit/loss in dollar.

**The floating leg** is referenced to the short-term interest rate. For most of the floating rate instruments, the standard reference rate is the LIBOR rate (London Interbank Offered Rate). Unlike TBill rates, LIBOR is not based on actual transactions. Prior to September 2012, it was calculated by BBA (British Bankers' Association), compiled from quotes given by 16 major banks (eliminating the highest and lowest our bank quotes and then averaging the remaining eight). Effectively, the 3M LIBOR is an indication of the average rate a leading bank can obtain unsecured funding for three months. It is a vital benchmark interest rate to which hundreds of trillions of dollars of financial contracts are tied, including CME Eurodollar futures and interest rate swaps. Figure 2 plots the time-series of the spread between the 3M LIBOR and 3M TBill rates. As you can see, the LIBOR spread tends to spike up during financial crises, reflecting the increased concerns on banks' credit quality.

The scandal on LIBOR fixing is an event that reflects very poorly on Wall Street. Even before the collapse of Lehman, questions about the accuracy of LIBOR showed up in articles published in *FT*, *WSJ*, and others. As we now know, LIBOR manipulation has been perceived as business as usual and were even encouraged in some banks. As of now, regulators in the US, UK, and EU have fined banks more than $9 billion for rigging LIBOR. The scandal cost the Chairman and CEO of Barclays their respective jobs, sent one trader to jail on a fourteen-year sentence, and the trials and investigations are still ongoing.

Coming back to the floating leg, in a standard swap contract, the floating leg is referenced to the 3M LIBOR rate and resets every quarter. For example, at time 0, we enter into a swap. The floating leg takes the prevailing 3M LIBOR rate at that time, $r_0^{3M}$, as a reference rate. Three months later, the floating payment is $r_0^{3M}/4\times$ notional amount. Now the prevailing 3M LIBOR rate becomes $r_{0.25}^{3M}$, which will be used as the

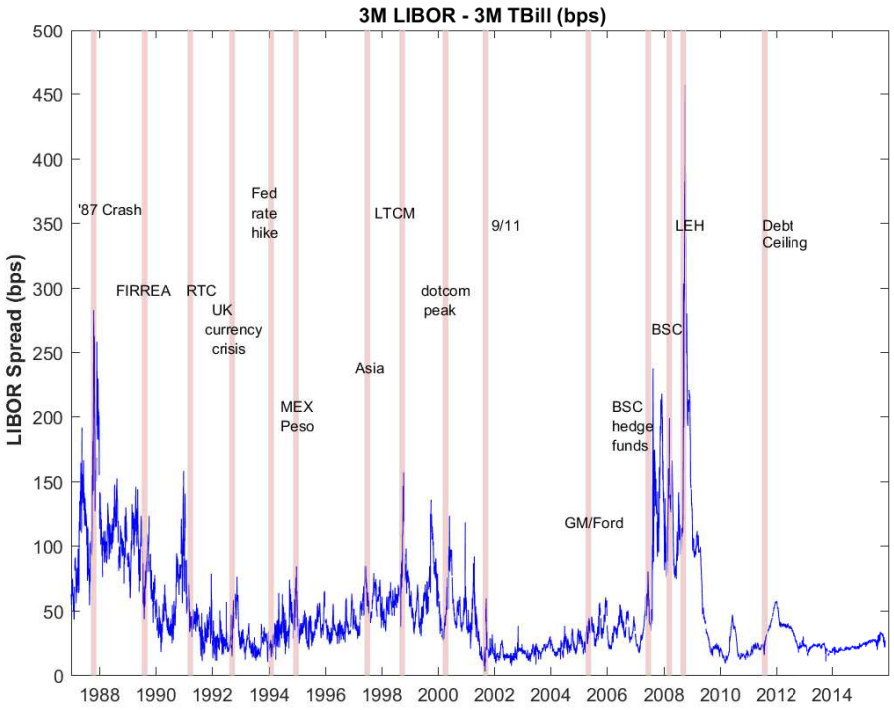


Figure 2: Time Series of the Spread between the 3M LIBOR and 3M TBill Rates.

reference rate for the next floating payment: $r^{3M}_{0.25}/4\times$ notional amount. As you will see, this feature, often called payment in arrears, is very important design for interest rate swaps.

- **Interest Rate Swap Rates:** A swap is structured so that the present value of the future cashflow equals to zero for both counterparties. As a result, the both counterparties enter into a swap without having to put up any capital initially. This is true in theory. In practice, because of concerns over counterparty credit risk, upfront payments are required for some counterparties for some type of contracts (e.g., credit default swaps).

  At the time 0, when two counterparties enter into an interest rate swap, the swap rate $s$ is set so that the present value of the future cashflow equals to zero for both counterparties. Once set, this swap rate will be fixed to the contract until its maturity in year $T$. To see how we can set this swap rate $s$, let's add the notional amount, say \$100, back to both the fixed and floating legs. Remember, this notional amount never exchanges hand, but adding it back serves a purpose for our understanding.

  Now, the fixed leg looks exactly like a Treasury bond of maturity $T$. The only difference is that instead of paying at the Treasury par-coupon rate $c$, the payer is paying the swap rate $s$. So the present value of this fixed cashflow is the same as the market price of a bond paying at a coupon rate of $s$. And what is the present value of the floating leg? For a bank whose three-month financing rate is the 3M LIBOR rate, it is ... \$100. The present value of these two cashflows should equal. So the present value of the fixed leg should equal to the present value of the floating leg, which is \$100. In other words, $s$ should be the coupon rate of a par coupon bond.

- **Duration:** It is really as simple as this. If you know how to do bond math for coupon-bearing bonds, you should be quickly at home with calculating swap rates. Because entering into a five-year interest rate swap as a receiver is the same as buying a five-year par-coupon bond and selling a five-year floating rate bond.

  As soon as we enter into the swap, the real effect on our portfolio is our exposure to interest-ate risk. By buying a five-year par-coupon bond, we load up on duration. For example, the modified duration for a five-year par-coupon bond is about 4.74 when the interest rate is at 2%. By selling the five-year floating rate, what is your interest rate exposure? Very little. The initial duration is 3 months, decreases to zero just before the next quarterly settlement, and reset to 3 months at the next quarterly settlement.

So if you would like to buy duration, you enter into an interest swap as a receiver; if you would like to sell duration, you enter as a payer.

Thinking in terms of duration is the most effectively way to understand this product. In fact, duration is the reason of existence for interest rate swaps.

- **Swap Spreads:** By now, it is obvious that interest rate swaps parallel Treasury bonds. In fact, the Wall Street Journal used to publish two curves everyday, as shown in Figure 3.


Figure 3: Treasury and Swap Curves, April 2007.

Figure 4 plots the time-series of the Swaps rates and Treasury rates. As you can see, these two markets are closely connected. In fact, it is a common practice to calculate the spread of swap minus Treasury rates of the same maturity and look at the swap spread, which is plotted in Figure 5.

The time-variation of the swap spreads is a very interesting topic, which we will not have time to talk too much about. The negative swap spreads at the longer maturity,

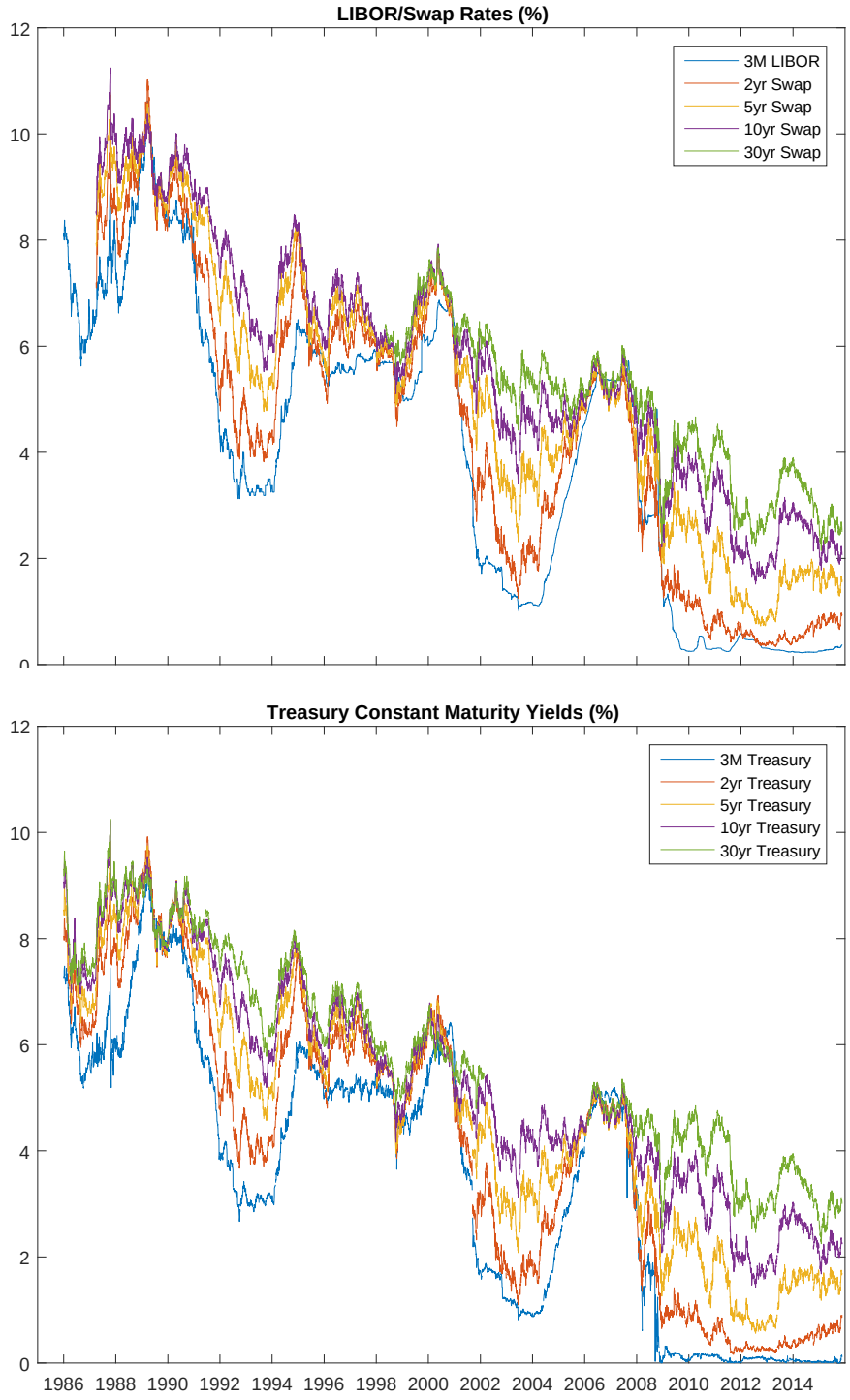


Figure 4: Time-Series of Treasury and Swap Rates.

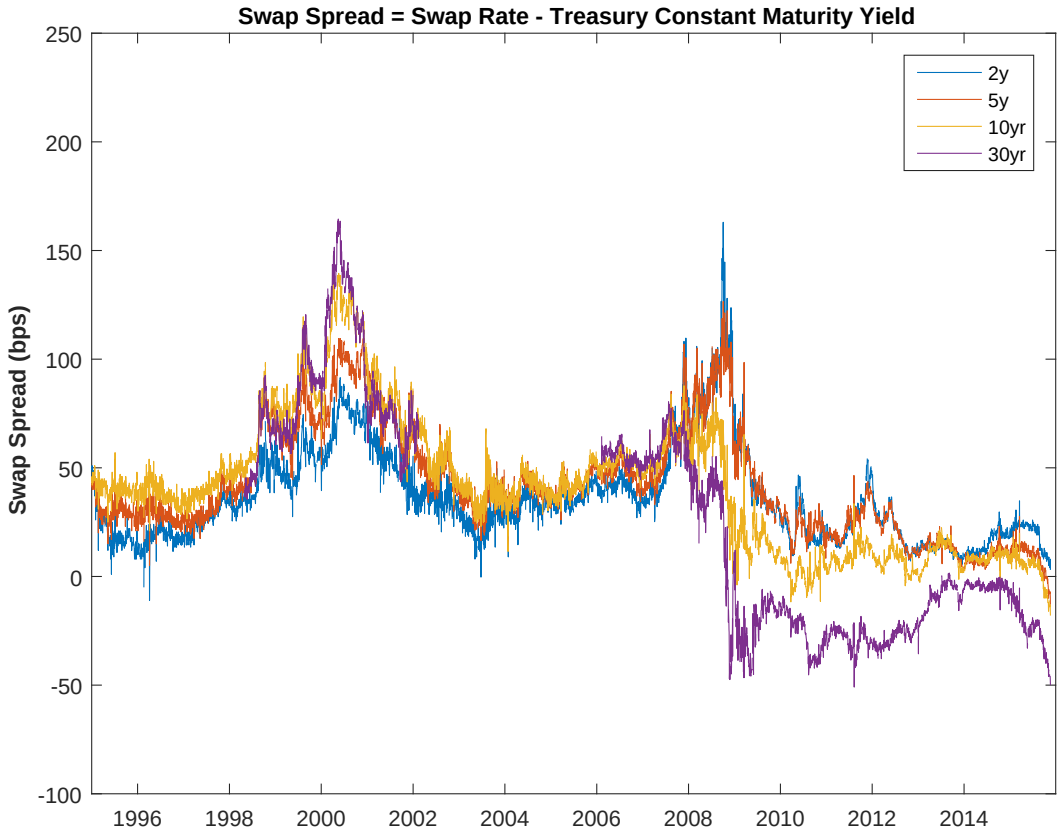


Figure 5: Time-Series of Treasury and Swap Rates.

especially for the 30yr, has been an interesting phenomenon after the Lehman default. If you asked me back in 1995, can spread spreads can negative, I would say very unlikely. But the 30yr swap spreads has been in the negative territory for the past 7 years.

# 2 Using Interest Rate Swaps

As discussed earlier, interest rate swaps are vehicles for duration. We enter into an interest rate swap as a receiver to buy duration, and as a payer to sell duration. To emphasize on this point, let's look at a specific example.

- **Negative Convexity of MBS:** Because of the prepayment options given to homeowners and mortgage borrowers, mortgage-backed securities (MBS) have negative convexity. Let me quote Fannie in its 2010 10K report,

  > Our mortgage assets consist mainly of single-family fixed-rate mortgage loans that give borrowers the option to prepay at any time before the scheduled maturity date or continue paying until the stated maturity. Given this prepayment option held by the borrower, we are exposed to uncertainty as to when or at what rate prepayments will occur, which affects the length of time our mortgage assets will remain outstanding and the timing of the cash flows related to these assets. This prepayment uncertainty results in a potential mismatch between the timing of receipt of cash flows related to our assets and the timing of payment of cash flows related to our liabilities.
  >
  > Changes in interest rates, as well as other factors, influence mortgage prepayment rates and duration and also affect the value of our mortgage assets. When interest rates decrease, prepayment rates on fixed-rate mortgages generally accelerate because borrowers usually can pay off their existing mortgages and refinance at lower rates. Accelerated prepayment rates have the effect of shortening the duration and average life of the fixed-rate mortgage assets we hold in our portfolio. In a declining interest rate environment, existing mortgage assets held in our portfolio tend to increase in value or price because these mortgages are likely to have higher interest rates than new mortgages, which are being originated at the then-current lower interest rates. Conversely, when interest rates increase, prepayment rates generally

slow, which extends the duration and average life of our mortgage assets and results in a decrease in value.

Effectively, when interest rates fall, mortgage borrowers take advantage of the prepayment option by refinancing their mortgage. As a result of this refinancing activity, some of the mortgage loans with nice looking duration (e.g., 30-year or 15-year maturity) are suddenly turned into cash (with zero duration). And this happens when interest rate is falling, exactly when a bond holder expects duration to work in his favor. For the entire pool of mortgages, the net effect is that the average duration is a function of the probability of prepayment. With falling interest rates, the probability of prepayment increases and the expected average duration shortens; With increasing interest rates, the probability of prepayment decreases and the expected average duration lengthens.

Figure 6 plots the time-series of MBS rates and durations for the Barclays US MBS Index. As you can see, when rates decrease, there are sharp declines in MBS duration; when rates increase, MBS duration increases. As a result, the convexity of MBS is negative. Also plotted in Figure 6, is the time-series of yields and duration for Barclays 5-year Treasury index. As you can see, the duration increased slightly over time as the interest rate decreased after the 2008 crisis. This is the effect of positive convexity. The small periodic variation in duration was due to periodic rebalance of the index composition in order to main the maturity of the portfolio close to five years.

- **Hedging Interest Rate Risk at Fannie and Freddie:** For Fannie and Freddie, holing MBS implies positive exposure to duration risk as well as exposure to negative convexity. In order to manage the interest rate exposure, Fannie and Freddie issue debt (i.e., agency debt) that is a mixture of short- and long-term, non-callable debt and callable debt. The varied maturities and callability of their debt (i.e., liabilities) give them the flexibility to deal with the variation of duration (i.e., negative convexity) in their mortgage assets. At the day-to-day level, however, this flexibility on the issuance side does not match the more frequent change in duration on the asset side. For this reason, Fannie and Freddie use interest rate derivatives extensively for hedging purpose. For example, as of December 31, 2010, a hypothetical increase of 50 basis points in LIBOR rate cost Fannie $0.9 billion before derivatives and $0.2 billion after derivatives. In other words, using interest rate derivatives helped Fannie hedge out a large portfolio of its interest rate exposure. The benefit of hedging is more significant in 2014: before derivatives, a hypothetical increase of 50 basis points in LIBOR cost Fannie $1.9 billion before derivatives and $0.3 billion after derivatives.

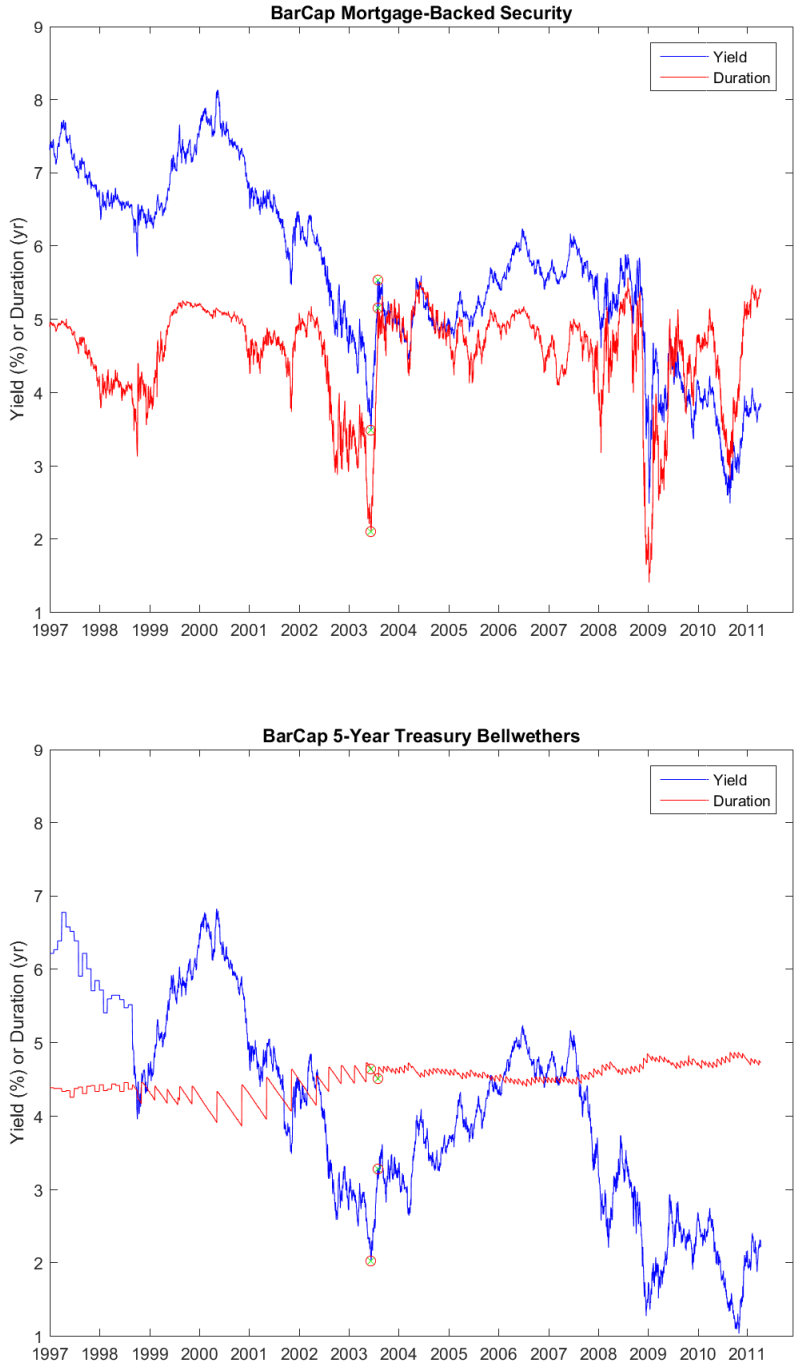


Figure 6: Time Series of Yield and Duration for Barclays US MBS Index and 5-Year Treasury Bellwethers Index.

| | Interest Rate Swaps | | | | Interest Rate Swaptions | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pay-Fixed | Receive-Fixed[2] | Basis[3] | Foreign Currency[4] | Pay-Fixed | Receive-Fixed | Interest Rate Caps | Futures | Other[5] | Total |
| | (Dollars in millions) | | | | | | | | | |
| Notional balance as of December 31, 2008 | $ 546,916 | $ 451,081 | $ 24,560 | $1,652 | $ 79,500 | $ 93,560 | $ 500 | $ — | $827 | $1,198,596 |
| Additions | 297,379 | 279,854 | 2,765 | 577 | 32,825 | 19,175 | 6,500 | — | 13 | 639,088 |
| Terminations[6] | (461,695) | (455,518) | (24,100) | (692) | (13,025) | (37,355) | — | — | (92) | (992,477) |
| Notional balance as of December 31, 2009 | $ 382,600 | $ 275,417 | $ 3,225 | $1,537 | $ 99,300 | $ 75,380 | $7,000 | $ — | $748 | $ 845,207 |
| Additions | 212,214 | 250,417 | 55 | 636 | 51,700 | 51,025 | — | 598 | — | 566,645 |
| Terminations[6] | (317,587) | (301,657) | (2,795) | (613) | (53,850) | (47,790) | — | (353) | (59) | (724,704) |
| Notional balance as of December 31, 2010 | $ 277,227 | $ 224,177 | $ 485 | $1,560 | $ 97,150 | $ 78,615 | $7,000 | $ 245 | $689 | $ 687,148 |
| Future maturities of notional amounts:[7] | | | | | | | | | | |
| Less than 1 year | $ 70,656 | $ 14,200 | $ 50 | $ 386 | $ 20,750 | $ — | $ — | $ 125 | $ 75 | $ 106,242 |
| 1 to less than 5 years | 90,788 | 168,000 | 35 | — | 35,300 | 4,500 | 7,000 | 120 | 593 | 306,336 |
| 5 to less than 10 years | 96,400 | 29,632 | 100 | 511 | 10,200 | 20,970 | — | — | 21 | 157,834 |
| 10 years and over | 19,383 | 12,345 | 300 | 663 | 30,900 | 53,145 | — | — | — | 116,736 |
| Total | $ 277,227 | $ 224,177 | $ 485 | $1,560 | $ 97,150 | $ 78,615 | $7,000 | $ 245 | $689 | $ 687,148 |

Figure 7: Risk Management Derivatives Held by Fannie Mae, from Fannie Mae's 2010 10K.

| | As of December 31,[2] | |
|---|---|---|
| | 2014 | 2013 |
| | (Dollars in billions) | |
| Rate level shock: | | |
| -100 basis points | $ 0.4 | $ 0.1 |
| -50 basis points | 0.1 | 0.0 |
| +50 basis points | (0.1) | (0.1) |
| +100 basis points | (0.1) | (0.5) |
| Rate slope shock: | | |
| -25 basis points (flattening) | 0.0 | 0.0 |
| +25 basis points (steepening) | (0.0) | 0.0 |

| | For the Three Months Ended December 31, 2014[3] | | |
|---|---|---|---|
| | Duration Gap | Rate Slope Shock 25 bps | Rate Level Shock 50 bps |
| | | Exposure | |
| | (In months) | (Dollars in billions) | |
| Average | 0.1 | $ 0.1 | $ 0.0 |
| Minimum | (0.3) | 0.0 | — |
| Maximum | 0.5 | 0.1 | 0.1 |
| Standard deviation | 0.2 | 0.0 | 0.0 |

Figure 8: Interest Rate Sensitivity of Net Portfolio to Changes in Interest Level and Slope of Yield Curve, from Fannie Mae's 2014 10K.

Figure 7 reports the derivatives used for hedging in Fannie's 2010 10K form. I am using the 2010 10K form in this plot because the reporting in Fannie's 2014 10K form is not as nice (in my personal opinion) as it was before. As of end 2010, Fannie's interest rate swap positions involve:

- Payer with notional amount $277 billion.
- Receiver with notional amount $274 billion.

In addition, Fannie also have positions on basis swaps, interest rate swaptions (options on swaps), interest rate caps (portfolio of options on floating interest rates, with each caplet providing the cap buyer the option to cap the short-term interest rate at a pre-arranged fixed level), and others.

Figure 8 also reports the overall interest rate sensitivity of Fannie's portfolio, including derivatives hedging, in 2014. As you can see, the average duration gap of Fannie was very small, only around 0.2 months, with a standard deviation around 0.2 months. The largest swings in duration gap were only in the range of -0.6 to 0.3 months. In addition to measuring its portfolio's sensitivity to interest rate (dollar duration and modified duration), Figure 8 also reports Fannie's sensitivity to change in the slope of the yield curve.

- **MBS Footprint on Swaps:** Going back to Figure 6, we see how changes in interest rate could affect the duration of MBS via the probability of prepayment. In June 2003, there was a sudden increase in interest rates. On June 13, 2003, the 10YR Treasury yield was at 3.13%, a result of a steady decline of interest rate. Over the next month and half, however, the 10yr rate increased steadily to 4.44% on August 1. As shown in Figure 6, the increase in the MBS rate was more dramatic, from 3.48% on June 13 to 5.53% on August 1, and the corresponding modified duration for the MBS increased from 2.09 to 5.15 years.

  This was an event that took the MBS market by surprise. With the increasing interest rate, the MBS holders found themselves to be long in duration through their MBS holdings. The liabilities side of their balance sheet does not match this sudden increase in duration. So they need to sell duration, fast. Otherwise, this unwanted duration will show up as a positive duration gap on their portfolio. Recall that the average duration gap for Fannie was 0.2 months in 2014. Selling duration means paying fixed in the interest rate swap market. In other words, selling swap "bonds." If this demand from MBS holders to sell duration arrives simultaneously in the swap market, how would

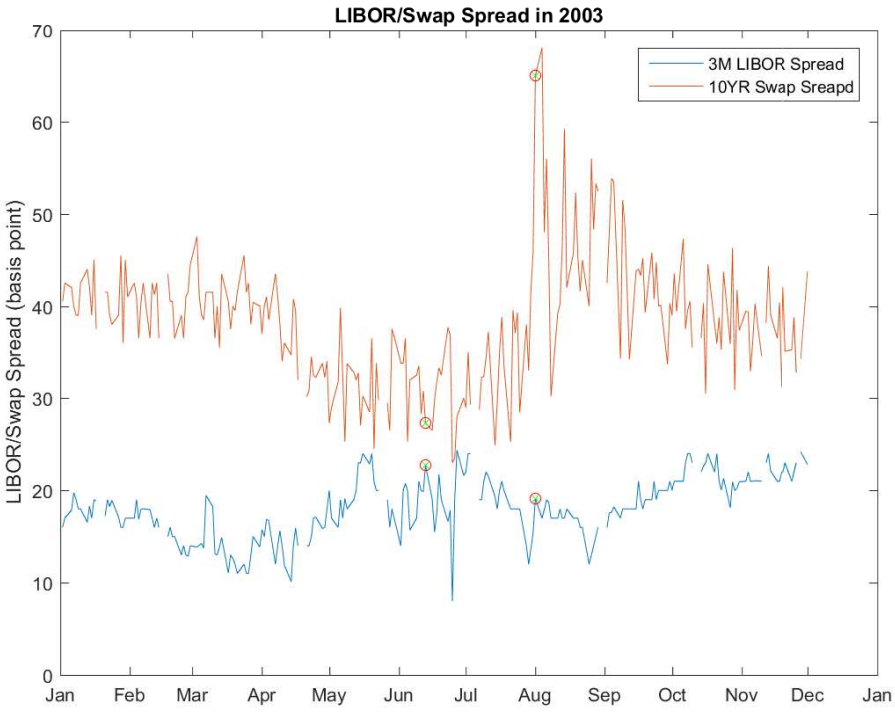


Figure 9: Time-Series of 3M LIBOR and 10YR Swap Rates in 2003.

it affect the price? The selling pressure on the swap "bond" would lead to decreasing "price" in the swap "bond" and increasing yield. As shown in Figure 9, this hedging activity resulted in a temporary spike in the 10yr swap spread just prior to August 1, 2013. For your viewing convenience, I've marked those two dates (June 13 and August 1) by a combined sign of green 'x' and red 'o'. Recall that the 10yr Treasury yield also increased during that period. So the increase in swap spread implies an increase in the 10yr swap rate that is more severe than the 10yr Treasury yield. By comparison, the 3M LIBOR spread was not affected during this period.

In going through this example of MBS hedging, we learn a few lessons. First, the negative convexity of MBS. Second, the wide usage of interest rate swaps as a hedging instrument. Third, the supply and demand in the swap market as a driver to the changes in swap spreads.

# 3 OTC Derivatives

- **An Overview:** The global OTC derivatives market had its beginning in the mid-1980s. Over the past 30 years, it has grown into an important part of the global financial markets, allowing business to manage and hedge financial risk. By far, the most important segment of this market is interest-rate product. As such, most of the hedging activities on interest rate risk have migrated from Treasury bonds to interest rate swaps. In addition, it also provides derivatives on currency, credit, equity, and commodities.

  This is a privately organized market, with transactions taking place bilaterally between dealers and end users, or dealers and other dealers. As such, the dealers function as market makers, engaged in either side of the transaction and keeping an inventory when necessary. The most active dealers are called the G14 dealers, comprising Bank of America-Merrill Lynch, Barclays Capital, BNP Paribas, Citi, Credit Suisse, Deutsche Bank, Goldman Sachs, HSBC, JP Morgan, Morgan Stanley, RBS, Societe Generale, UBS and Wells Fargo Bank.

  Because of the OTC nature of this market, trading information is very much limited. The lack of transparency in this market is in direct contrast to exchange-traded products such as equity, options, and futures, where trading information is readily available. For this market, the most comprehensive information available to the public is through the semi-annual and triennial surveys conducted by the Bank of International Settlements (BIS). This is a link to the most recent statistical release from BIS.

The OTC nature of this market also introduces potential systemic risk. Through their market-making activities in OTC derivatives, major dealers put themselves in the middle, creating an extra layer of interconnectedness in the financial system. At the time of its bankruptcy in September 2008, Lehman was a counterparty to 6,500 different institutions and corporations, across 1.2 million derivative transactions. The tangled web of deals took years to unwind and points to the obvious problem of this market.

Under Dodd-Frank, new rules are to be implemented with the objective of increasing transparency and reducing systemic risk in the derivatives markets:

  - Reporting swap transactions to a swap data repository;
  - Clearing sufficiently liquid and standardized swaps on central counterparties;
  - Where appropriate, trading standardized swaps on trading platforms; and
  - Setting higher capital and minimum margin requirements for uncleared swaps.

- **Notional Amount, Gross Market Value, and Credit Exposure:** After the 2008 crisis, the growth of this market has slowed down quite significantly. Figure 10 reports the total notional amount, surveyed semi-annually by the BIS, from 1998 through 2014. One potential contributing factor for the slowdown in the notional amount is compression. After the 2008 crisis and Dodd-Frank, some OTC trades were moved to CCPs (central clearing party), which facilitate the compression process. (Compression is a process for tearing up trades that allows economically redundant derivative trades to be terminated early without changing each participant's net position.)

  Figure 11 focuses on the more recent data. At end-June 2015, the total notional amount is at $553 trillion, the fourth consecutive semiannual decline of this market. As we see in this class, the notional amount of a derivatives contract is ... notional, never exchanging hands during the transaction.

  In order to measure the amounts at risk (or loss/profit), it is more useful to look at the gross market value of outstanding derivatives contract. At end-June 2015, the market was at $15.5 trillion, down from $20.9 trillion at end-2014 and $35 trillion at end-2008. The variation of gross market value is influenced by both the variation in notional amount as well as the variation in the underlying risk. A derivatives contract is structured so that, when the two counterparty enters into the contract, the gross market value equals to zero for both parties. Later, as the underlying risk fluctuates over time, the gross market value moves away from zero. The profit of one party equals

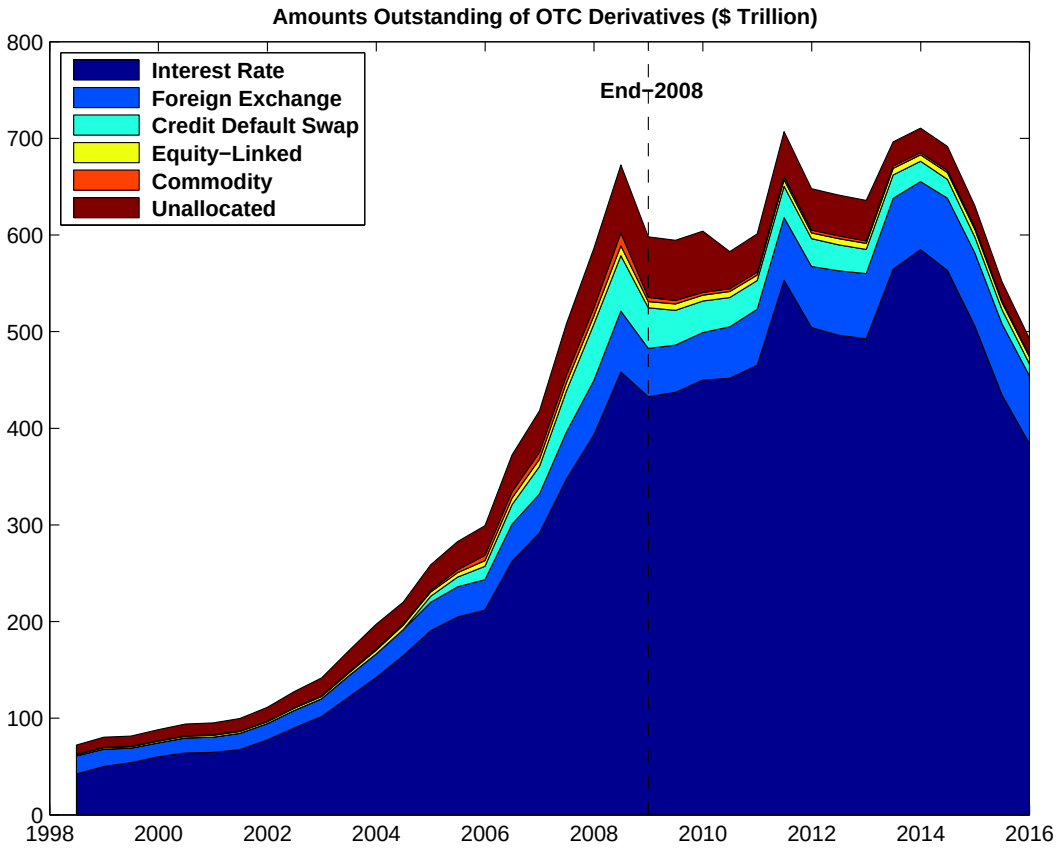


Figure 10: Global OTC Derivatives Markets, Notional Amount, Gross Market Value, and Gross Credit Exposure.

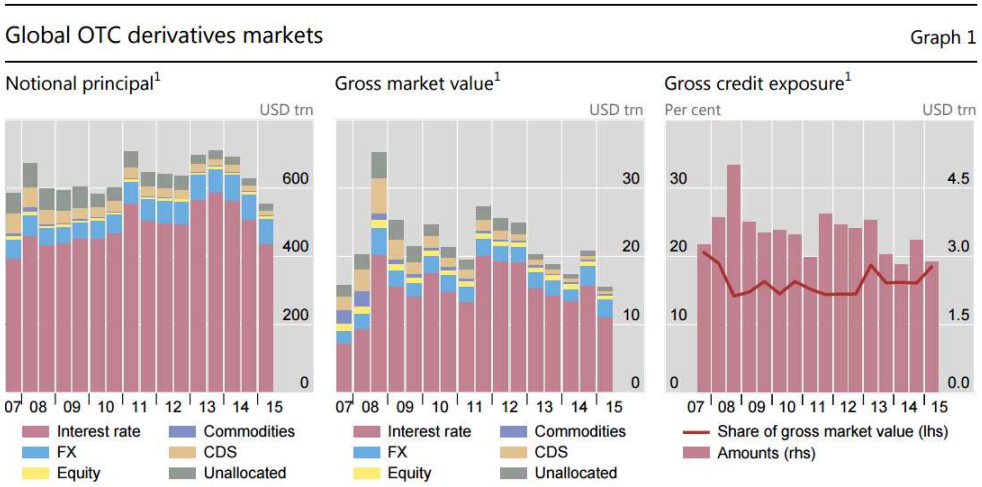


Figure 11: Global OTC Derivatives Markets, Notional Amount, Gross Market Value, and Gross Credit Exposure.

the loss of the other party, making the derivatives contract zero-sum as far as the two counterparties are involved.

If the underlying risk factor, say the interest rate, stays constant throughout the life of the contract, there the gross-market value stays at zero. If the gap between the initial interest rate and the current interest rate is close to zero, then the gross-market value will also be close to zero. In this sense, the gross-market value is also not the best way to capture the risk exposure. For the current example of interest rate exposure, a better measure of market risk exposure will be duration, as discussed early.

In addition to market risk exposure, counterparty credit exposure is also an important consideration in OTC derivatives. Again, this is due to the bilateral OTC nature of the market. Just imagine that you enter into a swap with Lehman and the gross market value of your position is positive and then Lehman went into bankruptcy. In order to reduce their exposure to counterparty credit risk, market participants can do two things. First, through netting agreements. For example, the last panel in Figure 11 reports the gross credit exposure, which adjusts the gross market value for legally enforceable bilateral netting agreements. With netting, the gross value of $15.5 turns into a net value of $2.9 trillion, accounting for 18.5% of the gross market value at end-June 2015.

Another way for market participants to reduce their counterparty exposure is through collateral. For example, suppose Goldman bought credit default swap (CDS) from AIG to insure against mortgage defaults. Initially, the contract was structured so that the present values are zero for both counterparties. After entering into the swap, the mortgage default started to increase, and Goldman's position is in the money. In a situation like this, Goldman would ask AIG to post more collateral against the gross market value of its CDS position. Often, the amount of the collateral is linked to the credit quality of the counterparty. A lower quality counterparty typically has to put up more collateral for the same amount of market exposure. This is why during the 2008 crisis, the downgrade of AIG was like throwing salt on the wound (although AIG totally deserved the treatment).

- **Goldman's Derivatives Book:** After the above discussions, we are now in a better position now to understand Figure 12, which was reported in Goldman's 10K report in 2014. First, it reports Goldman's derivatives positions by major product type on a gross basis. For example, the gross value of interest-rate derivatives totals to $786,362 million in assets and $739,607 million in liability with a total notional amount of

$47,112,518 million. In the language of derivatives, the assets are those derivatives on which Goldman is currently making money and the liability are those on which Goldman is losing money.

Just to compare with the broader market, as of December 2014, the total notional amount of interest-rate OTC derivatives was $505 trillion, making Goldman (total notional amount of $47 trillion) an important participant in this market.

To get to the $63 billion derivatives assets and $63 billion derivatives liabilities that eventually showed up in Goldman's financial statement as part of its financial instruments, counterparty netting was an important step. On the assets side, the gross value was $1,039,047 million and the counter party netting reduces it by $886,670 million, giving us $152,377 million, which corresponds to the gross credit exposure in the above discussion. Interestingly, $152,377 million accounts for 17.19% of the gross value $886,670 million, in line with the ratio reported in Figure 11 for the broader market in 2014.

The various components of the counterparty netting reported in Goldman's 10K is also interesting. In particular, "OTC-cleared" are done through central counterparties, which most likely is a product of Dodd-Frank. Also, the cash collateral netting also contributes to the reduction of the gross value to the net value used to measure counterparty exposure.

- **Interest Rate Swaps:** The interest rate segment accounts for the majority of the OTC derivatives activity. At end-June 2015, the notional amount of outstanding interest rate derivatives contract totaled $435 trillion, which represented 79% of the global OTC derivatives market. Within this segment, interest rate swaps, with total notional amount of $319 trillion, is by far the largest component. In term of market gross market value, interest rate swap was at $9.8 trillion by end-June 2015, a near 30% reduction from the gross market value of $13.9 trillion at end-2014. By comparison, the reduction in notional amount over the same period was 16%. As discussed earlier, the variation of gross market value is influenced by both the variation in notional amount as well as the variation in the underlying risk. The narrow gap between the initial interest rate and the current interest rate contributes to the low gross-market value.

  Figure 13 reports the interest rate derivatives by currency, maturity and counterparty. As we can see, the USD and Euro are the two major currencies for interest rate derivatives. Most of the swaps seem to be of less than five years in maturity. The longer maturity swaps accounts for less than 25% in 2015. The distribution of interest rate

| $ in millions | As of December 2014 Derivative Assets | Derivative Liabilities | Notional Amount |
|---|---|---|---|
| **Derivatives not accounted for as hedges** | | | |
| Interest rates | $ 786,362 | $739,607 | $47,112,518 |
| Exchange-traded | 228 | 238 | 3,151,865 |
| OTC-cleared | 351,801 | 330,298 | 30,408,636 |
| Bilateral OTC | 434,333 | 409,071 | 13,552,017 |
| Credit | 54,848 | 50,154 | 2,500,958 |
| OTC-cleared | 5,812 | 5,663 | 378,099 |
| Bilateral OTC | 49,036 | 44,491 | 2,122,859 |
| Currencies | 109,916 | 108,607 | 5,566,203 |
| Exchange-traded | 69 | 69 | 17,214 |
| OTC-cleared | 100 | 96 | 13,304 |
| Bilateral OTC | 109,747 | 108,442 | 5,535,685 |
| Commodities | 28,990 | 28,546 | 669,479 |
| Exchange-traded | 7,683 | 7,166 | 321,378 |
| OTC-cleared | 313 | 315 | 3,036 |
| Bilateral OTC | 20,994 | 21,065 | 345,065 |
| Equities | 58,931 | 58,649 | 1,525,495 |
| Exchange-traded | 9,592 | 9,636 | 541,711 |
| Bilateral OTC | 49,339 | 49,013 | 983,784 |
| Subtotal | 1,039,047 | 985,563 | 57,374,653 |
| **Derivatives accounted for as hedges** | | | |
| Interest rates | 14,272 | 262 | 126,498 |
| OTC-cleared | 2,713 | 228 | 31,109 |
| Bilateral OTC | 11,559 | 34 | 95,389 |
| Currencies | 125 | 16 | 9,636 |
| OTC-cleared | 12 | 3 | 1,205 |
| Bilateral OTC | 113 | 13 | 8,431 |
| Commodities | — | — | — |
| Exchange-traded | — | — | — |
| Bilateral OTC | — | — | — |
| Subtotal | 14,397 | 278 | 136,134 |
| **Gross fair value/notional amount of derivatives** | $1,053,444 [1] | $985,841 [1] | $57,510,787 |
| **Amounts that have been offset in the consolidated statements of financial condition** | | | |
| Counterparty netting | (886,670) | (886,670) | |
| Exchange-traded | (15,039) | (15,039) | |
| OTC-cleared | (335,792) | (335,792) | |
| Bilateral OTC | (535,839) | (535,839) | |
| Cash collateral netting | (103,504) | (36,155) | |
| OTC-cleared | (24,801) | (738) | |
| Bilateral OTC | (78,703) | (35,417) | |
| **Fair value included in financial instruments owned/ financial instruments sold, but not yet purchased** | $ 63,270 | $ 63,016 | |

Figure 12: Goldman's Derivatives Positions.

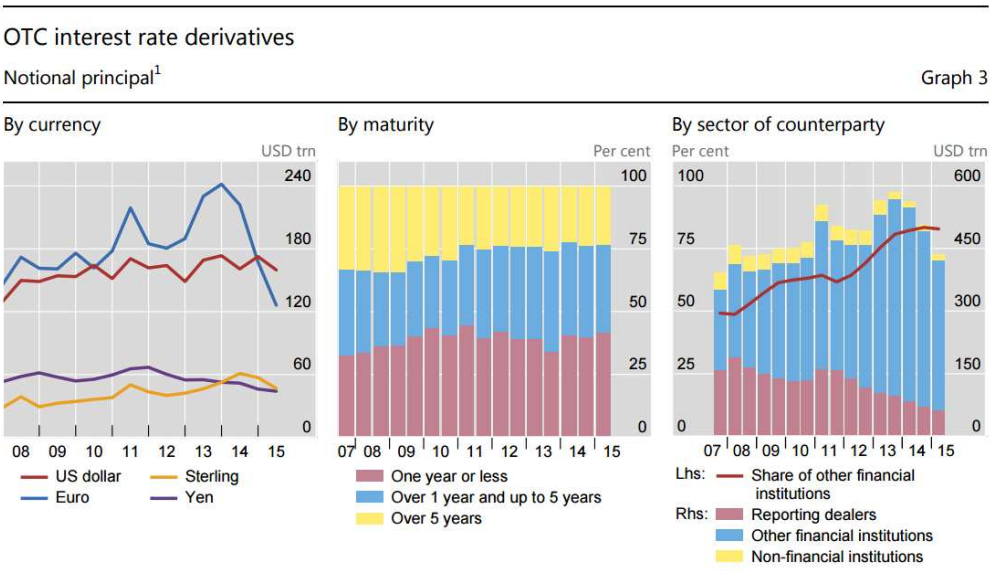


Figure 13: Interest Rate Derivatives by Currency, Maturity, and Counterparty.

derivatives by counterparty points to a continued shift in activity towards financial institutions other than dealers, including CCPs (central clearing party). It is also interesting to see that the need for non-financial institutions to hedge interest rate risk decreased significantly during the second half of 2013.

- **History of Swaps:** Although some swaps were arranged in the late 1970s, the first widely publicized swap took place in 1981 when IBM and the World Bank agreed to exchange interest payments on debt denominated in different currencies, an arrangement known as a currency swap. The first interest rate swap was a 1982 agreement in which the Student Loan Marketing Association (Sallie Mae) swapped the interest payments on an issue of intermediate-term, fixed-rate debt for floating-rate payments indexed to the three-month Treasury bill yield.

  Interest-rate swaps have existed since April 1987. In that time, these agreements between two parties to exchange periodic interest payments have risen from being an innovative means of transferring financial risk to providing a new "benchmark" for interest rates in the United States.

  Most swaps are entered with dealers, who then seek to limit their exposure to interest rate risk by entering into netting swaps with other counterparties. At end-June 2015, the notional amount of outstanding interest rate derivatives contract totaled $435 trillion, which represented 79% of the global OTC derivatives market. Within this

segment, interest rate swaps, with total notional amount of $319 trillion, is by far the largest component. In term of market gross market value, interest rate swap was at $9.8 trillion by end-June 2015.

In addition to swap dealers, major market participants include financial institutions and other corporations, international organizations such as the World Bank, government-sponsored enterprises, corporate bond and mortgage-backed securities dealers, and hedge funds. Hedging interest rate risk is one important motive for trading interest rate swaps. Swaps constitute the most common instrument in asset-liability management and in portfolio and debt management. A large universe of fixed-income securities including corporate bonds and mortgaged-back securities use interest rate swap spreads as a key benchmark for pricing and hedging. They are also widely used as a benchmark, as an index, and as an underlying asset for options (e.g. swaption).

Jun Pan
junpan@mit.edu

253-3083
E62-624

**Class 24: Fixed Income, Credit Risk**

This Version: December 1, 2016

# 1 The Credit Market

Banks play an important role in creating credit. In our class on Risk Management, we talked about a simple model of a bank, which holds only a fraction, say 10%, of its total deposits in its cash reserves and lends out the rest in loans. In doing so (often called fractional reserve banking), credit is created. Another important factor in credit creation is the Federal Reserve. In our class on Monetary Policy and Yield Curve, we talked about how the Fed injects or withdraws the total amount of bank reserves from the entire banking system via open market operations (i.e., purchasing or selling securities). In doing so, the Fed influences the amount of credit creation by banks. When the economy needs help, the Fed inject reserves into the system. With more reserves, banks are more willing to extend credit to their customers. When the economy is at risk of over-heating, the Fed withdraw reserves from the system and the banks are less willing to extend credit.

For anyone in Finance, understanding how the credit market works is essential. At the macro level, it informs you on the overall condition of the economy, given that the economic cycles are often linked to the credit market conditions. At the micro level, many of the financial instruments and much of the financial transactions contain credit components. So knowing how to model and price credit risk is an important skill to have.

Figure 1 plots the debt securities and loans outstanding in the US. By the second quarter of 2015, the total credit in the system is $62.30 trillion, among which $14.49 trillion goes to the federal government, $15.20 trillion goes to the financial sectors, $12.48 trillion goes to non-financial business and $14.05 trillion goes to households and non-profit organizations.

Figure 2 tracks the relative importance of the six sectors in Figure 1 by measuring their debt and loan outstanding as a fraction of the total credit. As you can, the fraction of borrowing from the financial sectors increased steadily from 12% in 1980 to a plateau of around 31% in the mid-2000 and decreased quite significantly to 24% in 2015. By contrast, the fraction of borrowing from the Federal Government increased quite dramatically from 11%

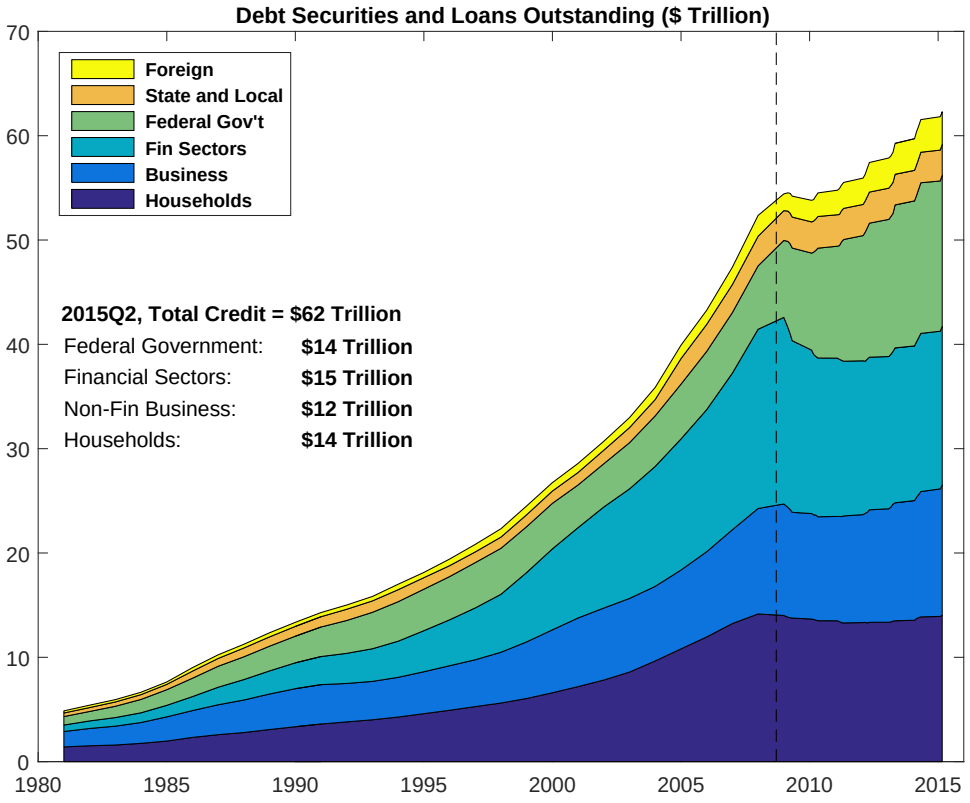


Figure 1: Debt Securities and Loans Outstanding.

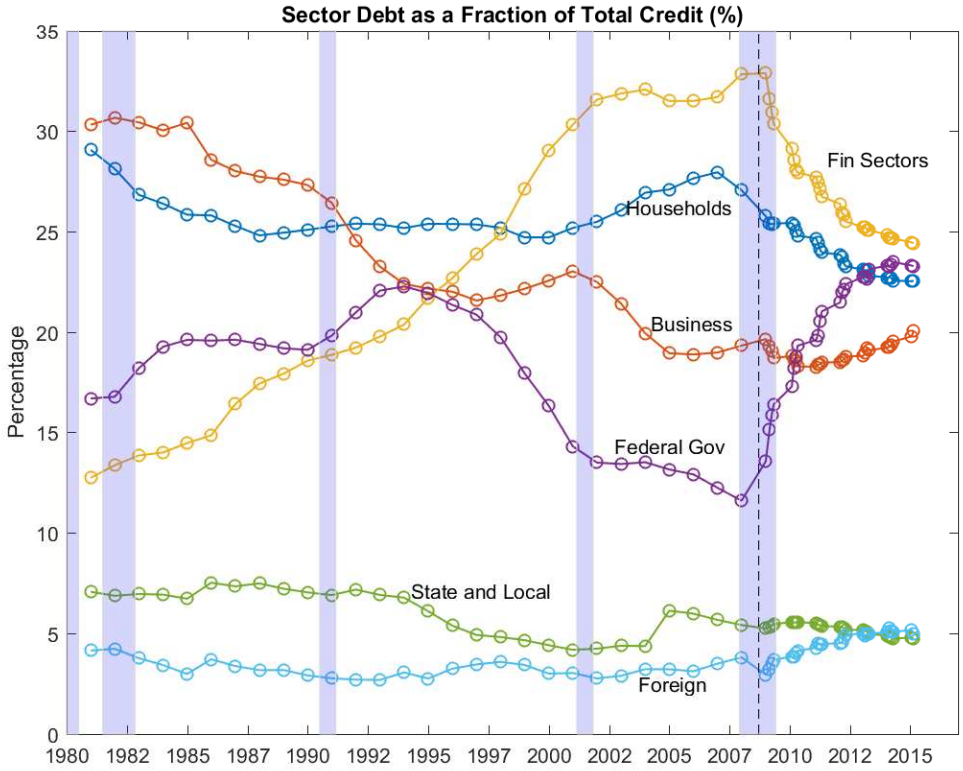


Figure 2: The Fraction of Debt Securities and Loans Outstanding by Sector.

before the 2008 crisis to 23% in 2015. Interestingly, the borrowing from non-financial business decreased from 30% in 1980 to about 19% before the 2008 crisis, dipped to an all-time low of 18.26% in the first quarter of 2011 and has recovered to 20% in 2015. For households, a large amount of borrowing is in mortgage. Compared with the federal government, non-financial business and financial sectors, the fraction of borrowing from the households remains relatively stable over time. This makes more obvious the significant increase and then the subsequent reduction in household borrowing surrounding the 2008 crisis. The fraction of households borrowing peaked at 27.9% in 2006 to dropped quite steadily to 22% in 2015.

# 2 Corporate Bonds

- **Default Probability:** In extending credit to a counterparty or an issuer, there are two considerations. First, the credit worthiness of the creditor. For this, we use the concept of *default intensity* to model the likelihood of the creditor's default. Second, in the event of a default, how much can we expect to recover or what is the expected loss? For this, we use the concept of *loss given default*. For the rest of the class, we will focus mostly on the corporate bond market, which by far is the most important component of the credit market.

  Figure 3 plots the annual issuer-weighted corporate default rates reported by Moody's in its "Annual Default Study: Corporate Default and Recovery Rates." This is a very useful document updated annually by Moody's and I would encourage you to take a look if you are interested in the credit market. To link the credit market condition to the business cycles, I've also plotted the NBER-dated recession periods in the same plot. The average annual default rate for investment-grade corporates is about 14 basis points. Excluding the great depression era, the average default rate is about 5.84 basis points. In general, the likelihood of default for an investment grade bond (the so-called fallen angle) is low. Since the great depression, the two largest investment-grade defaults were: the default of WorldCom in 2002 with $33B and the default of Lehman in 2008 with $120B.

  The annual default rates for speculative grades are significantly higher, with an average of 2.83% including the great depression era and 2.77% excluding the great depression. The connection of corporate defaults and business cycle is also stronger for this sector. For the three most recent recessions, the default rates all peaked above 10%, with the 2008 number reaching 13.3%.

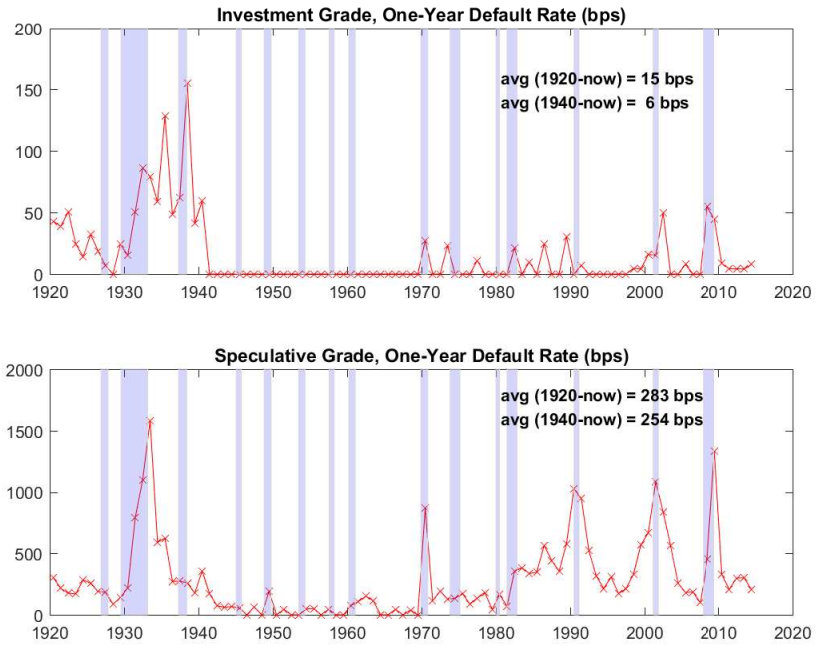


Figure 3: Annual Issuer-Weighted Corporate Default Rates for Investment and Speculative Grades.

- **Loss Given Default:** In the event of a default, it matters for the bond holder how much the bond is worth. For example, Fannie and Freddie had recovery rates close to 100 percent after receiving government guarantees on their senior unsecured bonds. Interestingly, the terms of the government rescues meant that "credit events" were deemed to have occurred, triggering CDS auctions. For most bonds, however, the recovery rates are much lower than 100%. For the purpose of credit pricing, the loss given default (LGD) is often set to a constant level around 50%.

  Figure 4 summarizes the average recovery rates reported by Moody's. For senior unsecured bonds (which is the category for most corporate bonds), the average recovery rate from 1982-2014 is about 37.4%, making the LGD around 1-37.4%=62.5%. In the most recent years (2013 and 2014), the recovery rates are higher compared with their long-term averages. This in part is due to the positive relation between higher recovery and economic conditions. For example, the recovery rate was 33.8% in 2008. It should be noted that measures such as recovery rates are sensitive to the specific default events. With defaults being rare in general, what we can learn from these default events is difficult to generalize. For example, in 2014, the recovery rate for senior sub-

EXHIBIT 7

**Average corporate debt recovery rates measured by post-default trading prices**

| | Issuer-weighted | | | Volume-weighted | | |
|---|---|---|---|---|---|---|
| Lien Position | 2014 | 2013 | 1982-2014 | 2014 | 2013 | 1982-2014 |
| 1st Lien Bank Loan | 78.4% | 75.1% | 66.6% | 80.6% | 67.7% | 62.5% |
| 2nd Lien Bank Loan* | 10.5% | 78.7% | 31.8% | 10.5% | 69.2% | 28.5% |
| Sr. Unsecured Bank Loan | n.a. | n.a. | 47.1% | n.a. | n.a. | 40.2% |
| Sr. Secured Bond | 59.5% | 59.8% | 52.8% | 76.5% | 59.5% | 52.4% |
| Sr. Unsecured Bond | 43.3% | 43.8% | 37.4% | 34.3% | 29.2% | 33.6% |
| Sr. Subordinated Bond* | 46.9% | 20.7% | 31.1% | 28.3% | 26.6% | 26.0% |
| Subordinated Bond** | 38.8% | 26.4% | 31.4% | 38.0% | 33.7% | 26.3% |
| Jr. Subordinated Bond | n.a. | n.a. | 24.7% | n.a. | n.a. | 17.1% |

* The average recovery rates for 2014's and 2013's second lien bank loans and senior subordinated bonds were each based on fewer than five defaults.

** The average recovery rates for 2014's subordinated bonds were based on fewer than five defaults.

Figure 4: Average Corporate Default Recovery Rates Measured by Post-Default Trading Prices.

ordinated bonds was 46.9%, slightly higher than the senior unsecured bonds (43.3%). This is due to the fact that the 46.9% number was based on only four defaults.

EXHIBIT 8

**Average corporate debt recovery rates measured by ultimate recoveries, 1987-2014**

| | Emergence Year | | | Default Year | | |
|---|---|---|---|---|---|---|
| Lien Position | 2014 | 2013 | 1987-2014 | 2014 | 2013 | 1987-2014 |
| Loans* | 81.0% | 76.7% | 80.2% | 68.4% | 76.6% | 80.2% |
| Senior Secured Bonds** | 57.1% | 84.2% | 63.0% | 59.4% | 56.9% | 63.0% |
| Senior Unsecured Bonds*** | 44.6% | 61.3% | 48.8% | 0.0% | 34.4% | 48.8% |
| Subordinated Bonds | 0.0% | 21.0% | 28.2% | 0.0% | 21.0% | 28.2% |

Figure 5: Average Corporate Default Recovery Rates Measured by Ultimate Recoveries.

The above recovery data are based on trading prices averaged over 30 days after a default event. An alternative recovery measure is based on ultimate recoveries, or the value creditors realize at the resolution of a default event. Going through the bankruptcy court is usually a lengthy process of 1-2 years following the initial default date. As shown in Figure 5, for senior unsecured bonds, the average ultimate recovery rate is 48.8% compared with the average recovery rate of 37.4%. In other words, funds specializing in distressed debt can purchase defaulted bonds and hope to recover the extra 10% by going through the bankruptcy court.

- **Estimating Default Probability:** We will now focus on modeling and estimating the default probability of a bond issuer, whiling keeping the recovery rate at a constant level.

Information about default probability can be collected from multiple sources. First, we can use credit ratings posted by the three rating agencies (Standard and Poor, Moody's, and Fitch). Using the historical default rates for each rating category, we can get a sense of the likelihood of default knowing the rating of the issuer. For example, as mentioned earlier, the average default rate of an investment-grade issuer is about 14 basis points including the great depression, and 5.84 basis points excluding the great depression. The corresponding numbers for a speculative-grade issuer are higher: 2.83% including the great depression era and 2.77% excluding.

In addition to measuring the default probability using the actual default experiences, there are two other markets from which we can collect information about default. First, a borrower with access to the corporate bond market is in general a large and mature company. In addition to debt financing, it also finance itself through the equity market. As such, information such as the firm fundamentals (e.g., financial statements) and equity market pricing and volatility becomes valuable information for us to assess the credit quality of the bond issuer. For this to work, however, we need to have a model that takes into account of the firm's total value (or cashflow) and prices the equity and bond securities simultaneously. These models are often called structural models of default, pioneered by the work of Merton (1974).

In addition to the equity market, the credit market itself provides a direct measure of default probability. We can use the bond yield spreads in the corporate bond market and the pricing of credit-default swaps to gauge the likelihood of default of an issuer.

The fact that the default probability of an issuer can be collected from these multiple sources implies that these markets are inter-related and any mis-pricing across the multiple markets could be an "arbitrage" opportunity. In fact, understanding the credit market requires a broad knowledge base that includes fixed-income, accounting, equity, and macro-economics. Once, a former MBA student visited my office with his boss and we spent a very nice hour discussing the CDS-bond basis. Toward the end, the boss said, with a full range of emotion, "When I met credits, that's when my love for Finance really started." Not that I wanted to stereotype middle-aged Wall-Street guys, but it was such an usual experience for me. And I fully understand what he meant.

- **Structural Model of Default:** This class of models are pioneered by the work of Merton (1974), and followed up and refined by Black and Cox (1976) and Leland (1994).

In Merton (1974), the total asset value $V$ of a firm is modeled as a geometric Brownian motion with growth rate $\mu$ and volatility $\sigma_A$:

$$dV_t = \mu\, V_t\, dt + \sigma_A\, V_t\, dB_t\,.$$

There are two classes of claimants for $V$: equity and bond holders. Let $K$ be the book value of the bond, which is a zero-coupon bond that maturities in some future date $T$. Let's consider the fixed time horizon, say $t$, which is prior to the maturity of the bond. In this model, as long as the time-$t$ firm value $V_t$ is above $K$, the firm is solvent. And default happens when $V_t$ falls below $K$.

Using what we learned from the Black-Scholes model, the distance to default can be measured by

$$DD = \frac{\ln(V/K) + (\mu - \sigma_A^2/2)\, t}{\sigma_A \sqrt{t}}$$

and the probability of default is $N(-DD)$. Although this is a very simple and somewhat unrealistic model, it captures the essence of what drives the probability of default for a name issuer. When the debt-to-asset ratio is high, the firm is closer to the default boundary $K$. As a result, its distance to default is small and the firm is more likely to default. When the firm's growth potential is good, then the the growth rate pulls the firm value away from the default boundary, making it less likely to default. A firm with more volatile asset value is more likely to touch the default boundary and therefore more likely to default.

In addition to probability of default, the Merton model can also price equity and bond simultaneously. The equity of a firm is essentially a call option of the firm's asset value $V$ with strike price $K$. A low leverage ratio $K/V$ implies the option is deep in the money. For example, the leverage ratio of an Aaa firm is around 13% while the leverage ratio of a Baa firm is around 43%. A typical single A-rated issuer has a leverage ratio around 30%. Even for a single B-rated issuer, the ratio of $K/V$ is around 65%, implying a call option that is deep-in-the-money. In the Merton model, the maturity of the call option is the maturity of the zero-coupon bond. But in practice, firms return to the capital market periodically to manage the maturity structure of their debt. This is where the Merton model (and many of the structural models of default) becomes inadequate.

On the bond side, buying a defaultable bond is the same as holding a default-free bond and selling a deep out-of-money put option on the firm's asset value $V$ with strike price

$K$.

- **Moody's KMV:** The Merton model was used by KMV to calculate expected default frequency (EDF). One of the key innovations of KMV was to recognize that, in the Merton model, the mapping between the distance to default to probability of default relies on the assumption of normal distribution: $N(-DD)$. Instead of using the mapping prescribed by the model, they use the actual default rates for companies in similar ranges to determine a mapping from DD to EDF. Effectively, they are using an empirical mapping. The EDF service provided by KMV has been quite successful and was acquired in 2002 by Moody's in a $210 million cash transaction.

- **Model Default using Reduced-Form Approach:** Let $\tilde{T}$ be the random default time of a credit issuer. Let $t$ be the horizon over which we care about the survival of this issuer. If $\tilde{T} \geq t$, then issuer is able to survive over the time horizon of our interest and the probability of survival can be summarized by Prob $\left(\tilde{T} \geq t\right)$. Conversely, the probability of default before time $t$ is $1 - \text{Prob}\left(\tilde{T} \geq t\right)$.

  Let's model this random default time $\tilde{T}$ by exponential distribution:

$$\text{Prob}\left(\tilde{T} \geq t\right) = e^{-\lambda t}$$

  , where the constant parameter $\lambda$ captures the default intensity. An issuer with large $\lambda$ defaults faster. To see this, let's consider the one-year default rate under this model. Setting $t$ to one year, the one-year survival probability is $e^{-\lambda}$ and the one-year default probability is $1 - e^{-\lambda} \approx= \lambda$, where I used the linear approximation for $e^x$ for small $x$. As you can see from this exercise, $\lambda$ is directly linked to default probability.

  We can now use this model to price defaultable bonds. I am going to side step the question about risk-neutral pricing for now. Let's consider a one-year zero-coupon bond with a face value of $1 and assume that the loss given default is 100%. Let $r$ be the riskfree interest rate (continuously compounded), the present value of the defaultable bond is:

$$P = e^{-r}\text{Prob}\left(\tilde{T} > 1\right) = e^{-r} \times e^{-\lambda} = e^{-(r+\lambda)}$$

  So the yield to maturity of this defaultable bond is $r + \lambda$ and the credit spread is $\lambda$.

  If the loss given default is not 100%, then, assuming the loss given default is $L$, we

have

$$P = e^{-r}\text{Prob}\left(\tilde{T} > 1\right) + e^{-r} \times \text{Prob}\left(\tilde{T} \leq 1\right) \times (1 - L) = e^{-(r+\lambda)} + e^{-r}\left(1 - e^{-\lambda}\right) \times (1 - L)\,,$$

where the second term comes from the recovery rate (1-L). For small $\lambda$, the credit spread can be approximated by $\lambda \times L$. Being able to recover part of the bond value makes the credit spread smaller. For two issuers with the same default intensity $\lambda$, the one with higher loss rate is priced with a higher credit spread. For example, an issuer might issue two bonds with different seniority, say senior vs. subordinated. Then the difference in the pricing of these two bonds bolts down to the recovery rate.

- **Historical Default Rates and Credit Spreads:** The reduced-form approach gives us a useful tool to connect the historical default rates to credit spreads. In Figure 6, I plot the Moody's one-year default rates together with the credit spreads of Barclays' investment-grade and speculative-grade bond indices. In calculating the credit spreads, I use Barclay's Treasury bond index. For teaching purpose, this is Okay, but the maturity match between the credit indices and the Treasury index are not very well done.

  In the first panel of Figure 6, the blue line plots the credit spreads for investment grades and the average is around 150 basis points. The red line plots the one-year default rates and the average over the same sample period is about 9 basis points. Recall that the one-year default probability is $1 - e^{-\lambda} \approx \lambda$ for small $\lambda$. Let the red line is essentially $\lambda$ for investment grade issuer and the average default intensity is 9 basis points. Also recall that the credit spread can be approximated by $\lambda \times L$, for small $\lambda$. So the blue line is essentially $\lambda \times L$. If we start with the red line, and multiply it by $L$, how can we get the blue line?

  Of course, in doing this calculation, we assume a constant default intensity and we make no distinction between risk-neutral and actual pricing. In other words, there is no role of default risk premium in our very simple pricing framework. In the academic research, this disconnect between the actual default experiences and the credit spreads has been studied quite extensively. It is called the "credit spread puzzle." The main puzzle is that the credit spreads are too high (or corporate bonds are too cheap) compared with the actual default experiences. Possible explanations include: credit risk premium and liquidity premium.

  In the second panel of Figure 6, the same exercise is done for speculative grades. There, the disconnect is not as severe. The average credit spread is 570 basis points while

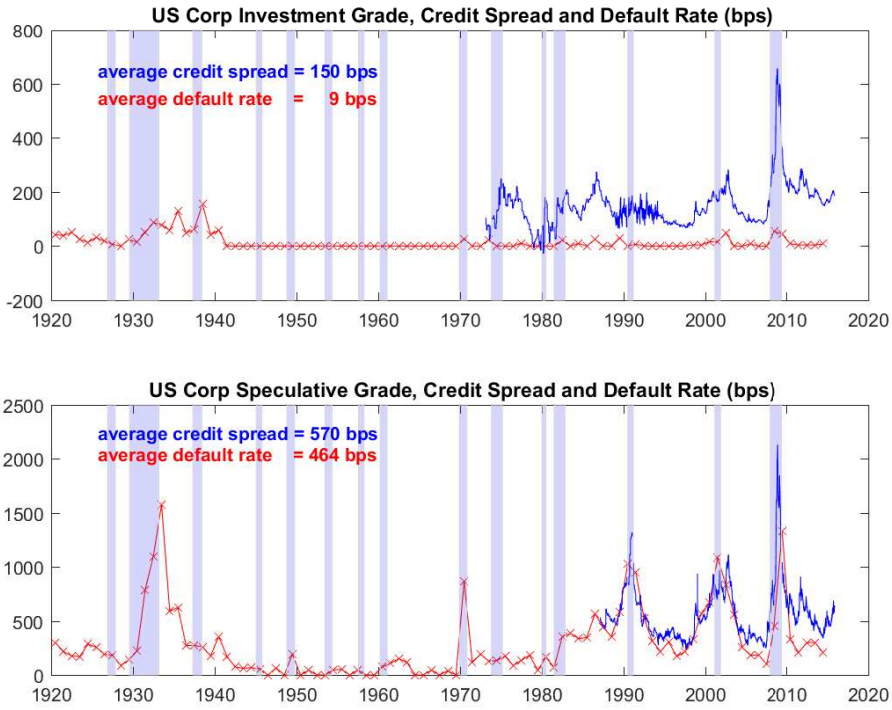


Figure 6: Corporate Default Rates and Credit Spreads for Investment and Speculative Grades.

the average one-year default rate for the same sample period is 464 basis points. If we start with the red line, and multiply it by $L = 1 - 37.4\% = 62.6\%$ to get the credit spread, we will still have a gap. In other words, even for the speculative grade, the credit spreads are too high. But the gap is not as dramatic as that for the investment grade.

# 3 Credit Default Swaps

- **Introduction:** The US corporate bond market is among the most illiquid markets. For a market of \$8T in 2015, the average daily trading volume is only \$25B. By comparison, the average daily volume is \$499B for US Treasury and \$321B for US Equity.

  In buying a corporate bond, investors take on both duration and credit exposures. To have a pure positive exposure to credit risk, investors have to hedge out the duration risk. To have a pure negative exposure to credit risk, investors have to locate, borrow, and then sell the bonds and buy back the duration exposure. The emergence of credit derivatives was in part a response to the limitations of corporate bonds as a vehicle for credit risk.

  Figure 7 plots the size of the CDS market along with the market for interest rate swaps. The market for CDS started out in the late 1990s. It really took off in the mid-2000 and peaked to a notional amount of \$58T in 2007. The notional amount of CDS has declined quite rapidly in recent years and is at \$16T in 2014.

- **CDS:** An investor enters into a CDS contract to either buy or sell credit protection on a named issuer, who could be a corporate issuer (e.g., Ford, GM, etc) or a Sovereign issuer (e.g., Russia, Mexico, etc). In 2009, the CDS market has gone through a pretty large change, which is called the "CDS big bang."

  Let me first describe the contract specification in the "old-fashioned" way. The swap has two legs: the fixed leg consists of quarterly fixed payments indexed to the CDS spread/price and the floating leg pays nothing as long as the named issuer is not in default. In the event of a default (before the maturity of the CDS contract), the payment of the fixed leg stops and the floating leg pays the full face value of the defaulted bond minus the recovery of the bond. In other words, in the event of a default, the seller of the protection makes the bond whole for the buyer.

  When the CDS was first introduced, the settlement involves the buyer locating and delivering the physical bond to the seller in exchange for the face value of the bond.

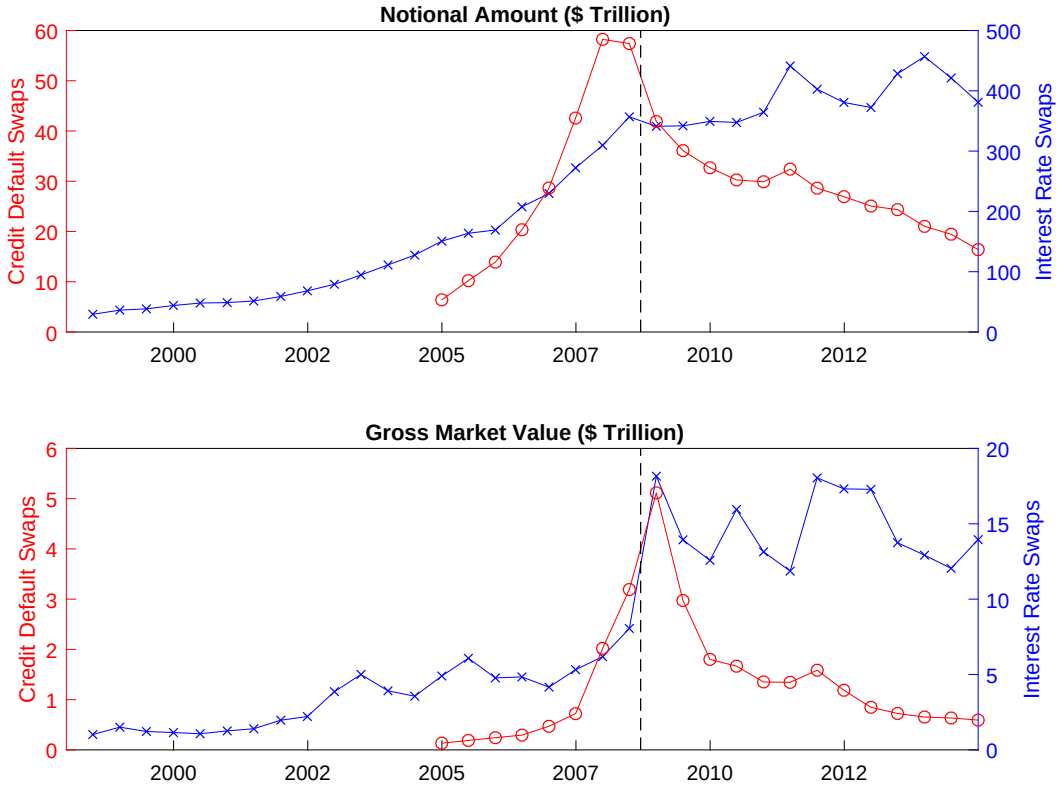


Figure 7: Interest Rate Swaps and Credit Default Swaps, Notional Amount and Gross Market Value.

In some default cases, however, the amount outstanding of CDS contracts became so large that it had the potential to drive up the price of the defaulted bonds as investors scrambled to acquire bonds to deliver. Because of this concern, in some cases, auctions can take place to determine the final recovery rate of a defaulted entity.

As with any swap, the present value of the fixed leg must equal the present value of the floating leg at the start of the contract and this is how the CDS spread/price is determined. Let's think of a very simple example to get some intuition. Suppose there is a one-year CDS on a named issuer. The fixed leg pays only annually. So the present value of the fixed leg (annuity):

$$\text{CDS} \times P\left(\tilde{T} > 1\right) \times e^{-r}$$

And the present value of the floating leg (insurance protection):

$$\text{Loss} \times P\left(\tilde{T} \leq 1\right) \times e^{-r}$$

We set CDS so that the two legs have the same present value:

$$\text{CDS} = \frac{P\left(\tilde{T} \leq 1\right) \times \text{Loss}}{1 - P\left(\tilde{T} \leq 1\right)}$$

As you can see, for small one-year probability of default $P\left(\tilde{T} \leq 1\right)$, the CDS can be approximated by

$$\text{CDS} \approx P\left(\tilde{T} \leq 1\right) \times \text{Loss} = \text{1yr Default Rate} \times \text{Loss}$$

Now let's apply the constant default intensity model to the above calculation:

$$\text{one-year default probability} = 1 - e^{-\lambda}$$

So the one-year CDS price is

$$\text{CDS} = \frac{\left(1 - e^{-\lambda}\right) \times \text{Loss}}{e^{-\lambda}} \approx \lambda \times \text{Loss}\,,$$

where the approximation works for small $\lambda$.

- **A Few Examples:** By now, it should be clear that the CDS price/spread is simply

the credit spread of a name issuer. For a corporate bond, we can invert its yield from its price. To determine the credit quality, we have to locate a Treasury bond of similar maturity and subtract the Treasury yield from the corporate yield to get the credit spread. For example, during the time period leading up to the GM's default in 2009, the price of the GM bonds were actually going up because interest rates were going down. It is only after taking out the duration exposure and focusing on the credit spread, can we know the true credit quality of GM. By comparison, a CDS on GM gives us direct information about GM's credit quality. Figure 8 plots the time-series of CDS for a few corporate issuers. And Figure 9 plots the time-series of CDS for a few sovereign issuers.

- **The CDS Big Bang:** After the 2008 crisis, the CDS market has gone through some very important change. Let me focus on just one change in contract specification that I think you should know. As I mentioned earlier, a swap involves two counterparties. At the start of the swap, the present value is zero for both counterparties. One important change in the CDS market is such that this is no longer true. As of now, the contracts are still quoted in the "old-fashioned" way, but the actual contract specification is different. For a high-quality (investment grade) named issuer, the fixed leg of the CDS contract is indexed to 100 bps. If the actual credit spread is 150 bps for this named issuer, then the buyer of the protection has to pay upfront fee equaling the present value of the difference between 150 bps and 100 bps. For a low-quality (high yield/speculative grade) named issuer, the fixed leg is index to a fixed rate of 500 bps. Again, an upfront fee (or rebate) is made to adjust for the difference between the deal spread (i.e., 500 bps) and the actual credit spread. To be honest, I am surprised at this change, but it seems to be taking place in the market.

- **CDS-Bond Basis:** For the same issuer, we now have two credit spreads: one from the corporate bond market and the other from the CDS market. The difference between the two is called CDS-bond basis: CDS spread minus the bond yield spread. In theory, these two spreads should be close or the basis should be small.

  Figure 10 plots the basis during the 2008 financial crisis. During the depth of the crisis, the CDS-bond basis became very negative, to a level close to $-300$ bps on average. For some named issuers, the basis were as negative as $-500$ bps. A negative CDS-bond basis implies that, for the same named issuer, the bond yield spread is larger than the CDS spread. In other words, the cash bond is cheaper than the CDS "bond." An arbitrage trade on negative CDS-bond basis would involve buying the corporate bond,

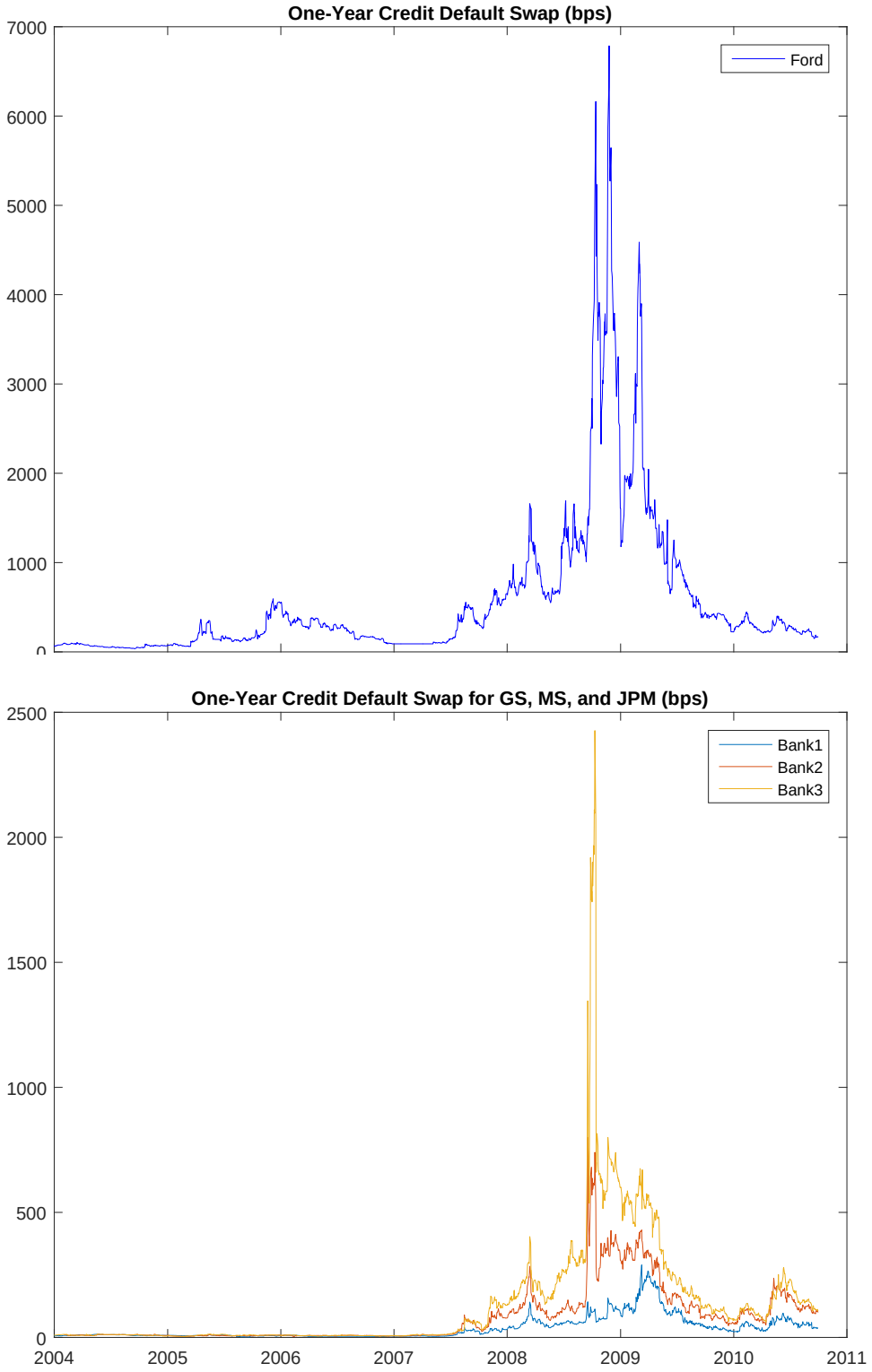


Figure 8: One-Year Credit Default Swaps on a few Corporate Issuers.


Figure 9: One-Year Credit Default Swaps on a few Sovereign Issuers.

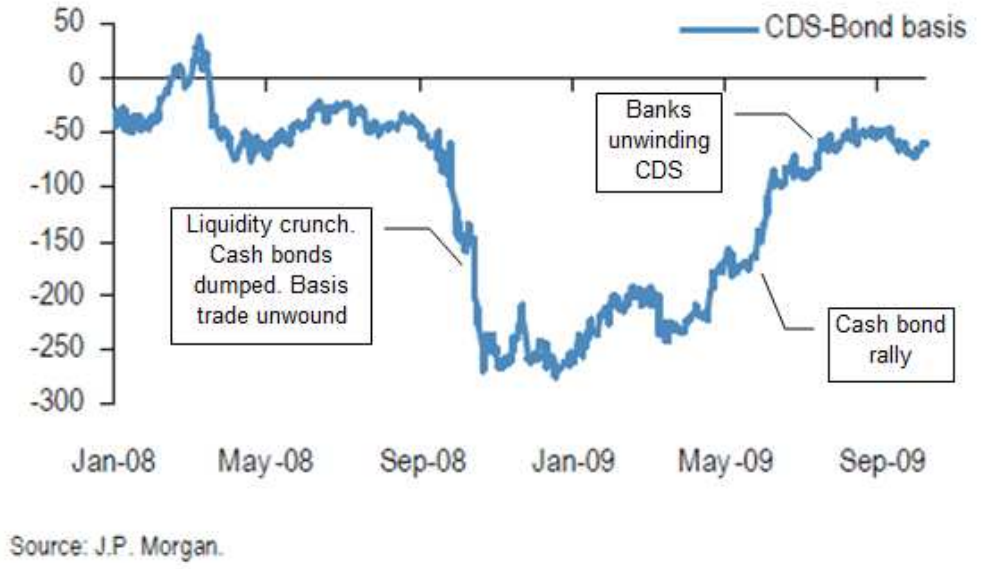


Figure 10: The CDS-Bond Basis in 2008-09.

hedging out the duration exposure and buying protection in the CDS market.

The evolution of the negative CDS-bond basis in 2008-09 was a story of limits to arbitrage. As shown in Figure 10, the basis turned negative after the Lehman default. According to the press, arbitrage trades were put on to bet that the negative basis would converge to zero. But instead of converging, the basis turned more negative. For example, Boaz Weinstein, a trader and co-head of credit trading at Deutsche Bank was down $1bn, Ken Griffin of Citadel was down 50% and John Thain of Merril was said to be down by more than $10bn. The big part of these losses was due to the "negative basis trade." As they unwound their negative basis trades due to losses, the basis further widened because of their unwinding. Eventually, the basis converged in the second half of 2009 as the financial markets recovered from the crisis.