

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

²See 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 μ is a number, while R_{t+1} is a random variable drawn from a distribution with mean μ and standard deviation σ .

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

estimate μ 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.

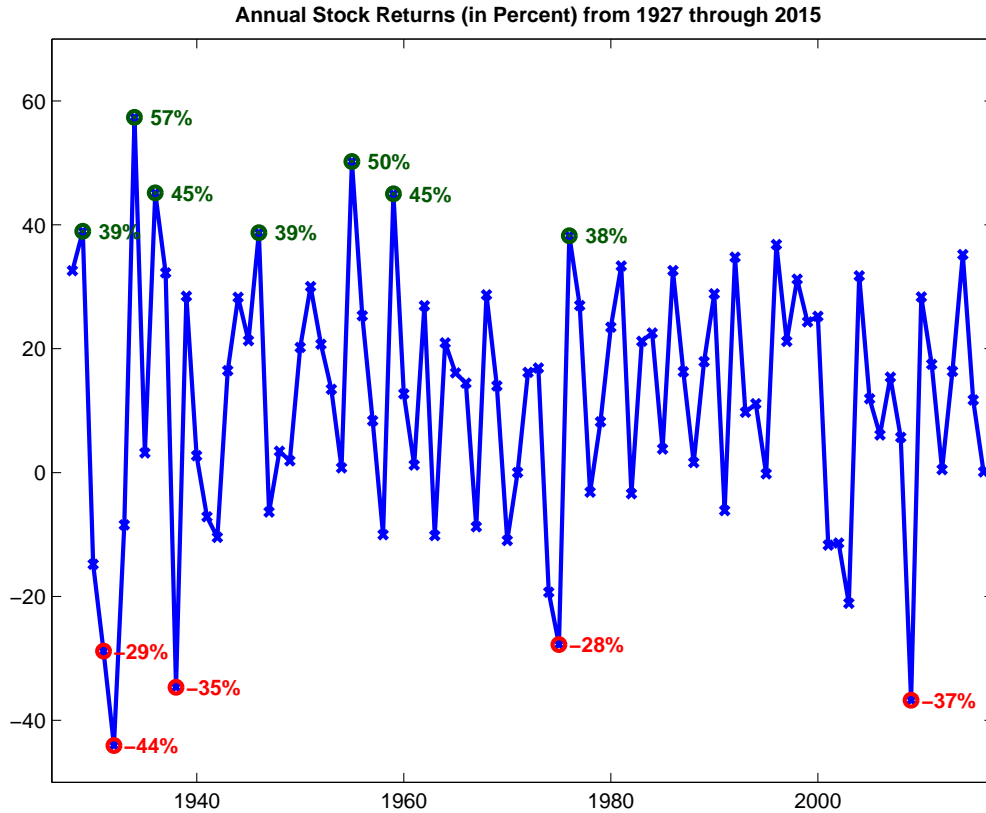


Figure 1: Time-series of annual stock market returns from 1927 through 2015. Returns are calculated using the CRSP-value weighted index, which includes all stocks traded on the US exchanges. Source: Prof. Ken French's website.

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 way that each R_t in the past was drawn from an identical distribution with mean μ and standard deviation σ . 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 μ and $\hat{\mu}$. Specifically, $\hat{\mu}$ is an estimator for an unknown number μ . 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 σ^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 μ is 12% and that of σ 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 σ). 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 μ . 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

³I'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 σ/\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%: $\text{normcdf}(-1.96)*2=0.05$. 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%.

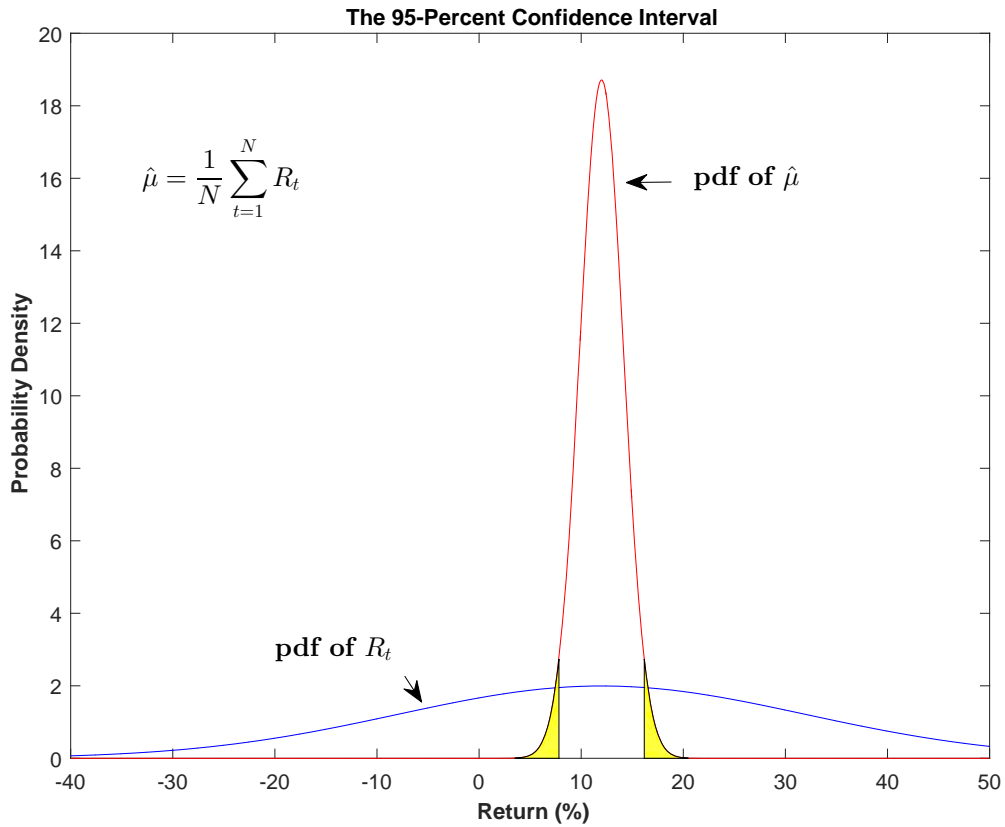


Figure 2: The probability density functions of R_t and $\hat{\mu}$ and the 95-percent confidence interval for the estimator $\hat{\mu}$.

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

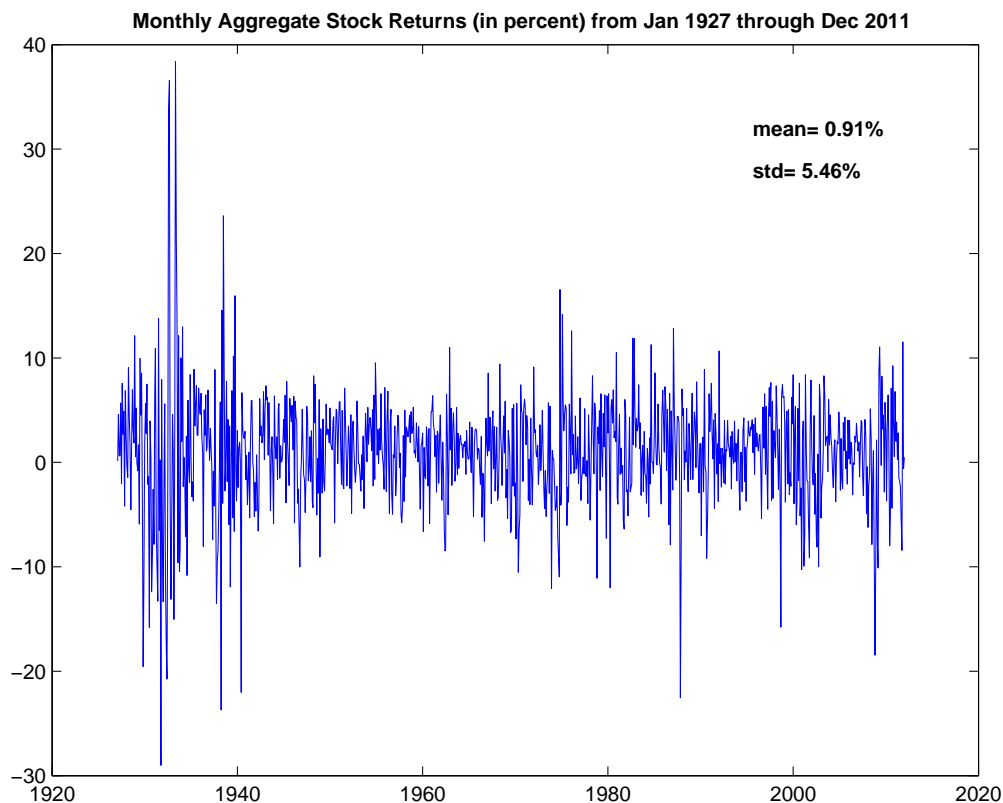


Figure 3: Time-Series of monthly stock returns from 1927 through 2011. Returns are calculated using the CRSP-value weighted index, which includes all stocks traded on the US exchanges. Source: Prof. Ken French’s Website.

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:

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

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

$$E(R^{\text{GE}}) - r_f = \beta^{\text{GE}} \times (E(R^M) - r_f) \quad (1)$$

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

$$R_t^{\text{GE}} - r_f = \alpha + \beta (R_t^M - r_f) + \epsilon_t. \quad (2)$$

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, ϵ_t . So there are three random variables involved R^{GE} , R^M , and ϵ , and two numbers to be estimated: α and β .

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 β is indeed the CAPM beta. This greatly simplifies the estimation process. Instead of doing $\text{cov}(R^{\text{GE}}, R^M)/\text{var}(R^M)$, we can simply run a regression to estimate β . Most importantly, the standard error of β 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 α , 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:

$$\alpha = R_t^{GE} - r_f - \beta (R_t^M - r_f) - \epsilon_t. \quad (3)$$

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

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

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

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

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 α 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 α , 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 α , we can now easily test whether or not α 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 β , because the volatility of r_f is much smaller than that of stock returns. But by dragging r_f along in this regression, we get α ! 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 α 's are significantly different from zero, then the CAPM, at 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).

⁴In fact, α is like μ for the time-series of $R_t^{GE} - r_f - \beta (R_t^M - r_f)$. So a good approximation for the standard error of α is $\text{std}(\epsilon_t)/\text{sqrt}(N)$, where ϵ 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 ϵ , which is the residual of the regression. If you like, you can back it out as a column in the Excel spreadsheet by,

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

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

Effectively, this regression decomposes R^{GE} into two random components: the first one is associated with the market portfolio through the term βR^M , and the other is the regression residual ϵ . 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:

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

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

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

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 μ 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 μ :** Since μ 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 μ . 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:

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

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 α , and can focus only on β :

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

- **The Regression Coefficient β :** 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:

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

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 β by running a regression.

- **The Standard Error:** Note that I've put a hat on β 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

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

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

Using σ_ϵ for the variance of the residual and σ_X for the variance of X (and remember that both Y and X have zero mean), we can further simplify the standard error to

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

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 σ_ϵ^2 has a direct impact on the precision of $\hat{\beta}$. Recall that the noisier the stock market (higher σ_R), the less precise the estimator $\hat{\mu}$. Here it is the ratio of σ_ϵ/σ_X that matters: for a given level of σ_X , the noisier the residual (the unexplained component), the less precise the regression coefficient β .

- **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:

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

Moreover, you will find the quite intuitive result of

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

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 α will not involve β . In this case, $\hat{\alpha}$ is indeed the mean estimator for the residual ϵ_t . Applying this to the CAPM setting, where σ_ϵ is the volatility of the idiosyncratic risk taken by a portfolio manager, you can see how the precision of his α is linked to his level of idiosyncratic risk. Having to estimate his β adds a bit noise to the precision of α , but it is the amount of the idiosyncratic risk that is the main driver for $\text{s.e.}(\hat{\alpha})$. So if a manager achieves his α through exposing his portfolio to high idiosyncratic risk, his signal to noise ratio will be low and the precision of his α 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 β could have very different R-squared's. You can dial up the idiosyncratic risk to decrease the R-squared while keeping the same β . 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

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

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 β (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 unbiasedness 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

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

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

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

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 (σ_ϵ) 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 β 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 β 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;

```