Stat 107: Introduction to Business and Financial Statistics

Class 15: Regression, Part II

Los Angeles Times

Op-Ed Sick of political polls? Try prediction markets

A hybrid between sports betting and derivatives markets, these allow traders to buy and sell shares that will pay out if a certain political

https://www.predictit.org/Browse/Group/67/National

Back to beta

- Let's look at Beta for LULU

Lululemon Athletica Inc. (LULU) ☆ Add to watchlist

NasdaqGS - NasdaqGS Real Time Price. Currency in USD

56.38 +1.50 (+2.74%)

As of 3:25 PM EDT. Market open.

Summary | Conversations | Statistics | Profile | Financials

Open	55.50	Market Cap	7.72B
Prev Close	54.88	P/E Ratio (ttm)	28.90
Bid	56.41 x 400	Beta	-0.00
Ask	56.42 x 300	Volume	2,480,817
Day's Range	55.14 - 56.79	Avg Vol (3m)	2,173,829
52wk Range	43.14 - 81.81	Dividend & Yield	N/A (N/A)
1y Target Est	71.00	Earnings Date	Dec 7, 2016 - Dec 12, 2016

Using R lets verify Beta

- Three years of monthly data

```
> fit=lm(luluret~spyret)
> coef(fit)
  (Intercept)         spyret
0.00009276783 0.01225231278
```

- What about a confidence interval?

```
> confint(fit)
                  2.5 %     97.5 %
(Intercept) -0.03747605 0.03766158
spyret      -1.16478361 1.18928823
```

Formally, The Market Model

- We can model the return on a stock (mutual fund, hedge fund, etc…) as

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

- where

α_i = part of the security's return independent of the market

R_M = rate of return on the market index

β_i = a constant that measures the expected change in R_i given a change in R_M

ε_i = randome noise, indepent of R_M (firm-specific surprises)

Statistical Properties of the MM

- What does the market model say about the variance of returns?
- According to the model

$$
\begin{aligned}
Var(R_i) &= Var(\alpha_i + \beta_i R_M + \varepsilon_i) \\
&= Var(\beta_i R_M + \varepsilon_i) \\
&= \beta_i^2 Var(R_M) + Var(\varepsilon_i) + 2\beta_i Cov(R_M, \varepsilon_i) \\
&= \beta_i^2 Var(R_M) + Var(\varepsilon_i)
\end{aligned}
$$

Add some new notation

- We have

$$Var(R_i) = \beta_i^2 Var(R_M) + Var(\varepsilon_i)$$

- Which we will write as

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_{\varepsilon,i}^2$$

Decompose the Risk of Security i

- We can decompose the risk (variance) for security *i*:

$$\sigma_i^2 = \underline{\beta_i^2 \sigma_M^2} + \sigma_{\varepsilon,i}^2$$

Part of security i's variance due to market news and movement-**called market** or **systematic risk or non-diversifiable risk.**

Part of security i's variance due to non-market news (e.g. industry effects)-**called nonmarket or unsystematic risk or diversifiable risk.**

Interpretation of Beta

- Examine this formula one last time

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_{\varepsilon,i}^2$$

- Note that the market variance is the same for all securities.
- So, we say Beta measures a stock's degree of systematic or market risk.

Firm Specific versus Systematic Risks

- **Systematic** (market or macroeconomic) risk usually accounts for about 25% of a typical stock's total risk.
- **Firm-specific** risk usually accounts for about 75% of a typical firm's total risk.
- You can eliminate almost all the firm-specific risk by holding a well-diversified portfolio of stocks. The random bad and good firm-specific events are then largely offset within the portfolio.

Firm Specific versus Systematic Risk

- Most people would never put all of their wealth in one common stock.
  - Too much exposure to firm-specific risk
- However, many people would put a large amount of their wealth in a mutual fund that invests in a diversified portfolio of common stocks.
  - The relevant risk here is the systematic risk, as the firm-specific risk has been diversified away.
- **You can never diversify away systematic risk, as all stocks have exposure to this risk.**

The Beta for AAPL

- Run the regression (older data)

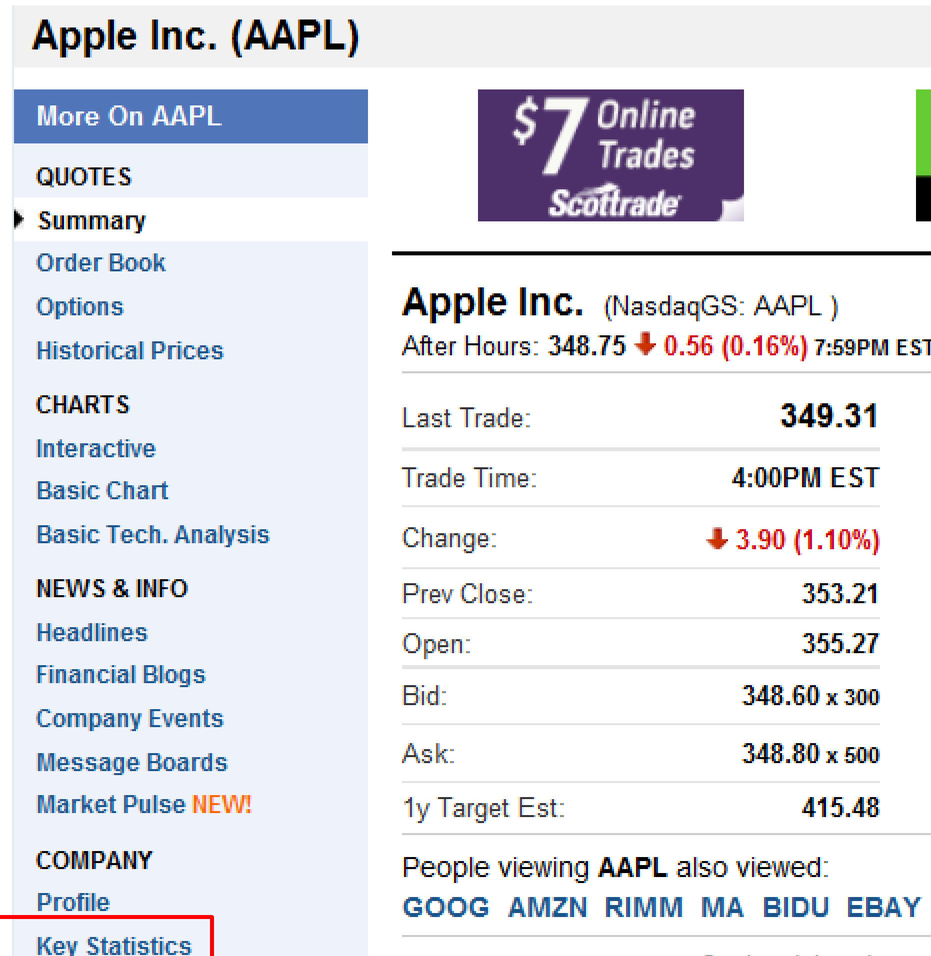
```
> fit=lm(aaplret~spyret)
> fit

Call:
lm(formula = aaplret ~ spyret)

Coefficients:
(Intercept)       spyret
    0.03225      1.20728
```

Compare with Yahoo

- Click on "Key Statistics"



AAPL Beta from Yahoo

Trading Information	
Stock Price History	
Beta:	1.20
52-Week Change[3]:	67.25%
S&P500 52-Week Change[3]:	16.81%
52-Week High (Feb 16, 2011)[3]:	364.90
52-Week Low (May 6, 2010)[3]:	199.25
50-Day Moving Average[3]:	346.42
200-Day Moving Average[3]:	309.23
Share Statistics	

We agree!
But note that Yahoo (and every web site for that matter) DOES NOT give a confidence interval for Beta. Why not?????

A Confidence Interval for Beta

- Its simply beta+/-2(standard error), but R can compute it automatically.

```
> confint(fit)
                  2.5 %     97.5 %
(Intercept) 0.007070099 0.05743848
spyret      0.795459482 1.61909834
```

- What does this imply??

Check on R^2

- According to the regression output, 50% of AAPL's variability is explained by market movements.

```
> summary(fit)

Call:
lm(formula = aaplret ~ spyret)

Residuals:
      Min        1Q    Median        3Q       Max
-0.241825 -0.038749  0.005398  0.044820  0.131521

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.03225    0.01241   2.600   0.0136 *
spyret       1.20728    0.20286   5.951 8.95e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07544 on 35 degrees of freedom
Multiple R-squared: 0.503,     Adjusted R-squared: 0.4888
F-statistic: 35.42 on 1 and 35 DF,  p-value: 8.947e-07
```

Let's find the Beta for JNJ

- JNJ (johnson and johnson), sort of a boring stock.

```
getSymbols("JNJ",from="2008-03-01")
getSymbols("SPY",from="2008-03-01")
spyret=monthlyReturn(SPY)
jnjret=monthlyReturn(JNJ)
```

Run the regression, compare to Yahoo

```
> fit=lm(jnjret~spyret)
> fit

Call:
lm(formula = jnjret ~ spyret)

Coefficients:
(Intercept)       spyret
 -0.0002453    0.5776404
```

Yup, we got this beta thing down I think

Trading Information	
Stock Price History	
Beta:	0.58
52-Week Change[3]:	-4.32%
S&P500 52-Week Change[3]:	16.81%
52-Week High (Apr 20, 2010)[3]:	66.20
52-Week Low (Jul 22, 2010)[3]:	56.86
50-Day Moving Average[3]:	61.17

Always give a confidence interval

- Estimates by themselves are always useless!
- Always report a confidence interval

```
> confint(fit)
                   2.5 %     97.5 %
(Intercept) -0.01136636 0.01087579
spyret       0.39578529 0.75949558
> 
```

Check on R^2

- Based on the regression output below, 54% of the variation of JNJ’s returns are related to movements in the market.

```
> summary(fit)

Call:
lm(formula = jnjret ~ spyret)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.070986 -0.018944  0.006906  0.019956  0.069626 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.0002453  0.0054781  -0.045    0.965    
spyret       0.5776404  0.0895791   6.448 1.99e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03331 on 35 degrees of freedom
Multiple R-squared: 0.543,     Adjusted R-squared: 0.5299 
F-statistic: 41.58 on 1 and 35 DF,  p-value: 1.991e-07
```

Portfolio Betas

- Consider an equally weighted portfolio of AAPL, JNJ, SBUX and MSFT.

```
> portret = (aaplret+jnjret+sbuxret+msftret)/4
> portbeta=coef(lm(portret~spyret))[2]
> portbeta
  spyret
1.005001
```

- It turns out that Betas add.

Betas Add

```
> aaplbeta=coef(lm(aaplret~spyret))[2]
> jnjbeta=coef(lm(jnjret~spyret))[2]
> msftbeta=coef(lm(msftret~spyret))[2]
> sbuxbeta=coef(lm(sbuxret~spyret))[2]
> (aaplbeta+jnjbeta+msftbeta+sbuxbeta)/4
  spyret
1.005001
> portbeta
  spyret
1.005001
```

Why Least Squares?

- As a quick aside, (least squares) regression fits a line by solving the equation

$$\min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

```
lm(y~x)
```

- Why not solve the following equation? This is called LAD (least absolute deviation) regression.

$$\min_{b_0, b_1} \sum_{i=1}^{n} | Y_i - b_0 - b_1 X_i |$$

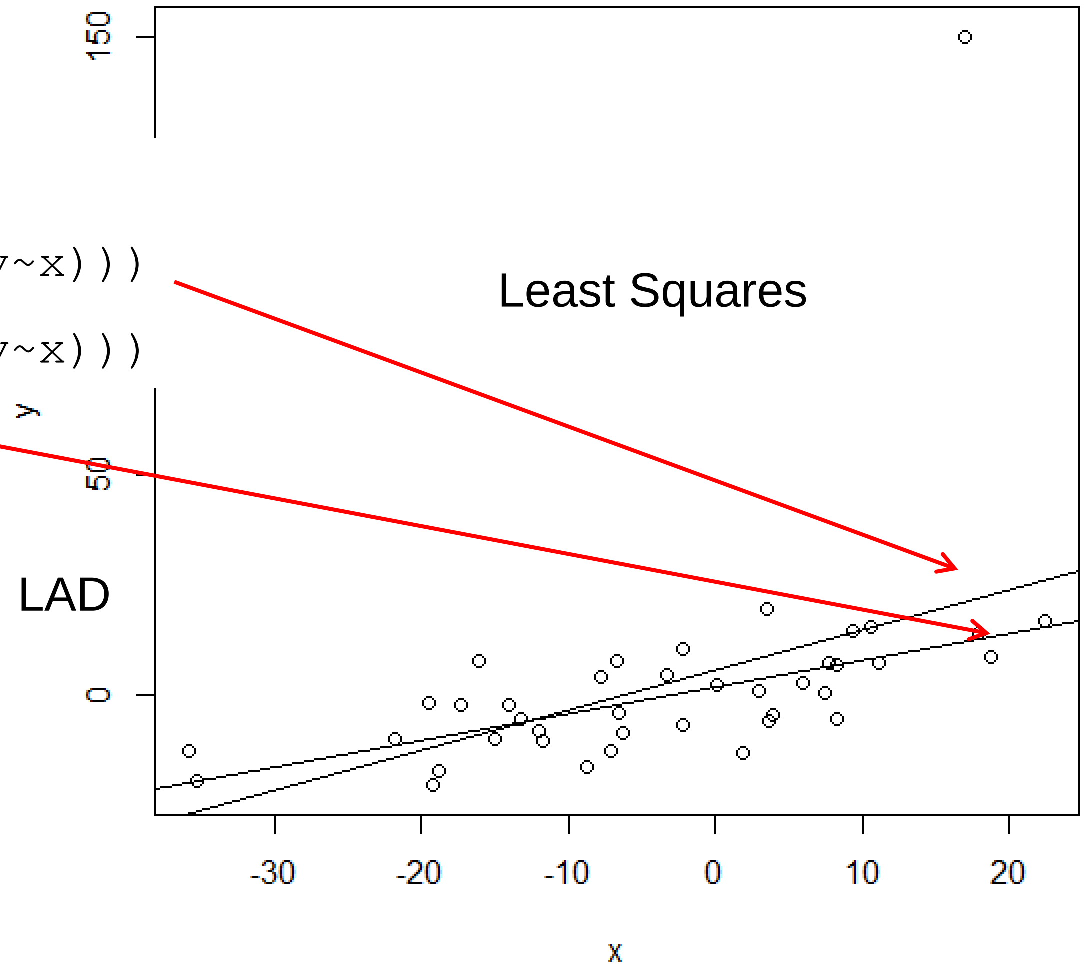
```
rq(y~x)
[in package
quantreg]
```

Least Squares versus MAD

```
plot(x,y)

lines(abline(lm(y~x)))

lines(abline(rq(y~x)))
```



LAD Betas

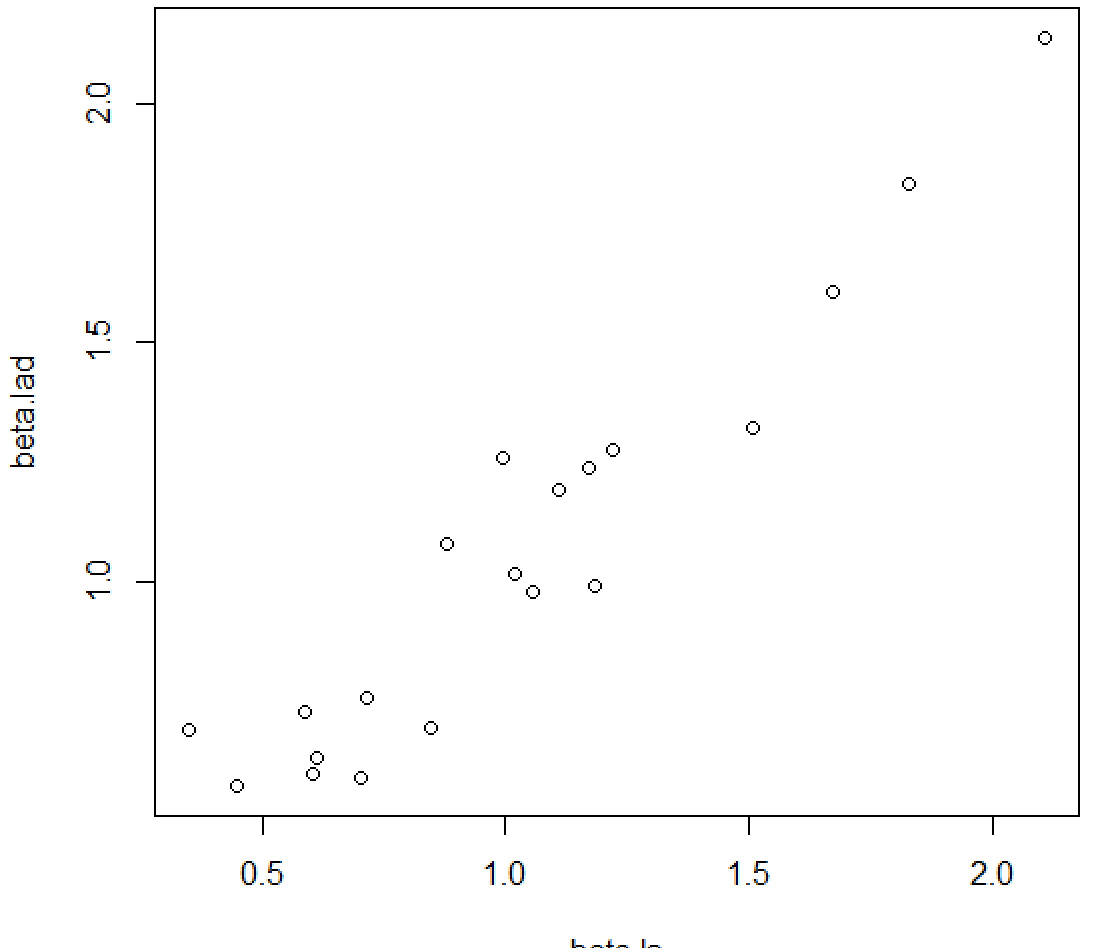
```
asset.names=c("ATVI", "ADBE", "AKAM", "ALTR", "AMZN",
"AMGN", "APOL", "AAPL", "AMAT", "ADSK", "ADP" , "BIDU",
"BBBY", "BIIB","BMC",  "BRCM" ,"CHRW" ,"CA" ,  "CELG",
"ANF")

n=length(asset.names)
getSymbols("SPY")
spy.ret=monthlyReturn(Ad(SPY))
beta.ls=1:n
beta.lad=1:n

for(i in 1:n){
      x=getSymbols(asset.names[i],auto.assign=FALSE)
       x.ret=monthlyReturn(Ad(x))
      beta.ls[i]=coef(lm(x.ret~spy.ret))[2]
      beta.lad[i]=coef(rq(x.ret~spy.ret))[2]
}
```

Which ones are correct?

```
> cor(beta.ls,beta.lad)
[1] 0.9531647
```



Beta is Beta until its not

- People love talking and thinking about Beta; it reduces the complexities of risk and comparing stocks to a single, easily interpretable number.
- However, Beta has numerous issues.
- Beta depends on the day of the week, the time scale, the market index being used, etc….

Some Issues with Beta

- Which market index?
- Which time intervals?
- Time length of data?
- Non-stationary
  - Beta estimates of a company change over time.
  - How useful is the beta you estimate now for thinking about the future?

The Beta for AAPL

YAHOO

Trading Information	
Stock Price History	
Beta:	1.18
52-Week Change[3]:	50.97%
S&P500 52-Week Change[3]:	11.37%
52-Week High (Feb 16, 2011)[3]:	364.90
52-Week Low (May 6, 2010)[3]:	199.25
50-Day Moving Average[3]:	349.39
200-Day Moving Average[3]:	319.52

Apple Inc. (NasdaqGS: AAPL)
After Hours: **341.00** ↓ 0.20 (0.06%) **5:44PM EDT**

Last Trade:	**341.20**	Day's Range:	**339.14 - 342.62**
Trade Time:	**4:00PM EDT**	52wk Range:	**199.25 - 364.90**
Change:	↑ 1.90 (0.56%)	Volume:	**11,639,820**
Prev Close:	**339.30**	Avg Vol (3m):	**17,462,300**
Open:	**342.58**	Market Cap:	**314.34B**
Bid:	**341.01 x 100**	P/E (ttm):	**19.04**
Ask:	**341.20 x 1000**	EPS (ttm):	**17.92**
1y Target Est:	**422.53**	Div & Yield:	**N/A (N/A)**

GOOGLE

Apple Inc. (Public, NASDAQ:AAPL) Watch this stock

341.20

+1.90 (0.56%)

Mar 22 - Close

NASDAQ real-time data - Disclaimer

Currency in USD

Range	339.14 - 342.62	P/E	19.05
52 week	199.25 - 364.90	Div/yield	-
Open	342.56	EPS	17.91
Vol / Avg.	11.64M/18.69M	Shares	921.28M
Mkt cap	314.34B	Beta	1.37

Reuters.com just confuses me

VALUATION RATIOS

	Company	Industry	Sector	S&P 500
P/E Ratio (TTM)	18.46	10.85	12.80	17.31
P/E High - Last 5 Yrs.	39.05	29.30	60.30	88.93
P/E Low - Last 5 Yrs.	18.91	10.50	11.61	12.27
Beta	1.39	1.18	1.04	1.32

Wtf? Maybe the average beta of all SP500 stocks??

Let's try GOOG

Trading Information	
Stock Price History	
Beta:	1.04
52-Week Change[3]:	5.01%
S&P500 52-Week Change[3]:	11.37%
52-Week High (Jan 19, 2011)[3]:	642.96
52-Week Low (Jul 6, 2010)[3]:	433.63
50-Day Moving Average[3]:	601.72
200-Day Moving Average[3]:	583.81

VALUATION RATIOS

	Company	Industry	Sector	S&P 500
P/E Ratio (TTM)	21.33	17.55	12.80	17.31
P/E High - Last 5 Yrs.	52.01	45.41	60.30	88.93
P/E Low - Last 5 Yrs.	22.58	12.55	11.61	12.27
Beta	1.20	0.81	1.04	1.32

Google Inc. (Public, NASDAQ:GOOG) Watch this stock

577.32

+0.82 (0.14%)

After Hours: **577.32** 0.00 (0.00%)

Mar 22, 6:19PM EDT

NASDAQ real-time data - Disclaimer

Currency in USD

Range	572.51 - 579.23	P/E	21.95
52 week	433.63 - 642.96	Div/yield	-
Open	577.27	EPS	26.30
Vol / Avg.	1.89M/2.85M	Shares	321.52M
Mkt cap	185.62B	Beta	1.19

What about JNJ?

Stock Price History	
Beta:	0.58
52-Week Change[3]:	-9.63%
S&P500 52-Week Change[3]:	11.37%
52-Week High (Apr 20, 2010)[3]:	66.20
52-Week Low (Jul 22, 2010)[3]:	56.86
50-Day Moving Average[3]:	60.24
200-Day Moving Average[3]:	61.92

VALUATION RATIOS

	Company	Industry	Sector	S&P 500
P/E Ratio (TTM)	12.26	21.69	45.51	17.31
P/E High - Last 5 Yrs.	18.55	27.43	58.69	88.93
P/E Low - Last 5 Yrs.	12.82	11.55	13.57	12.27
Beta	0.59	0.63	0.63	1.32

Johnson & Johnson (Public, NYSE:JNJ) Watch this stock

58.79

-0.04 (-0.07%)

After Hours: **58.81** +0.02 (0.03%)
Mar 22, 6:34PM EDT
NYSE real-time data - Disclaimer
Currency in USD

Range	58.70 - 59.15	P/E	12.31
52 week	56.86 - 66.20	Div/yield	0.54/3.67
Open	58.80	EPS	4.78
Vol / Avg.	10.16M/12.13M	Shares	2.74B
Mkt cap	161.10B	Beta	0.58

Beta's are Weird

- Consider the paper, "Are calculated betas worth for anything?"
- Great reading, particularly since Betas are ubiquitous and easily come up in finance discussions.

Table 1 from the paper

- Consider the following table, showing Betas collected from different web sites:

Table 1. Betas of different companies according to different sources

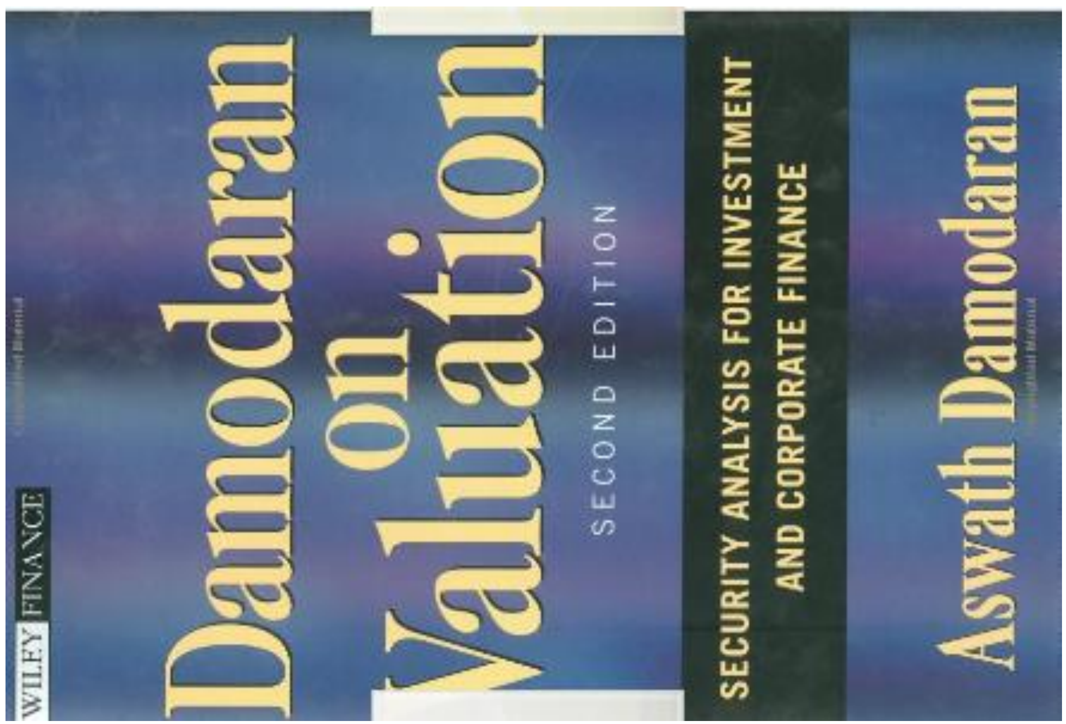
	AT&T	Boeing	CocaCola	Date
Yahoo	0.61	0.46	0.29	12-febr-03
Multex	0.87	0.66	0.42	12-febr-03
Quicken	1.14	0.66	0.41	12-febr-03
Reuters	0.87	0.68	0.42	12-febr-03
Bloomberg	1.00	1.07	0.64	12-febr-03
Datastream	1.10	1.10	0.37	12-febr-03
Buy&hold	0.84	0.66	0.41	14-febr-03

- Which one is correct?

Damodaran, 2001

Beta estimates for Cisco versus the S&P 500. *Source: Damodaran (2001, page 72)*

	Daily	Weekly	Monthly	Quarterly
2 years	1.72	1.74	1.82	2.7
5 years	1.63	1.70	1.45	1.78



Bruner (1998)

In addition to relying on historical data, use of this equation to estimate beta requires a number of practical compromises, each of which can materially affect the results. For instance, increasing the number of time periods used in the estimation may improve the statistical reliability of the estimate but risks the inclusion of stale, irrelevant information. Similarly, shortening the observation period from monthly to weekly, or even daily, increases the size of the sample but may yield observations that are not normally distributed and may introduce unwanted random noise. A third compromise involves choice of the market index. Theory dictates that R_m is the return on the market portfolio, an unobservable portfolio consisting of *all* risky assets, including human capital and other nontraded assets, in proportion to their importance in world wealth. Beta providers use a variety of stock market indices as proxies for the market portfolio on the argument that stock markets trade claims on a sufficiently wide array of assets to be adequate surrogates for the unobservable market portfolio.

Bruner (1998)

Exhibit 4. Compromises Underlying Beta Estimates and Their Effect on Estimated Betas of Sample Companies

	Bloomberg[a]	Value Line	Standard & Poor's
Number	102	260	60
Time Interval	wkly (2 yrs.)	wkly (5 yrs.)	mthly(5 yrs.)
Market Index Proxy	S&P 500	NYSE composite	S&P 500
Mean Beta	1.03	1.24	1.18
Median Beta	1.00	1.20	1.21

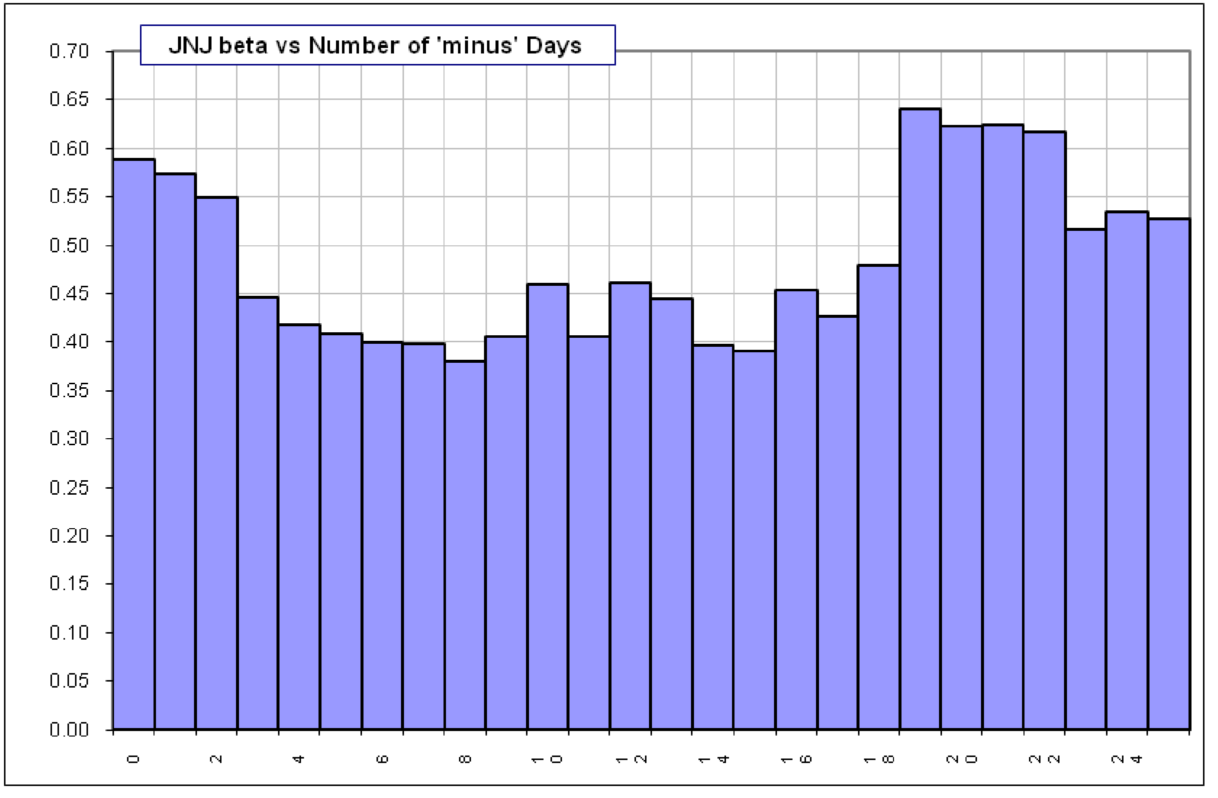
[a]With the Bloomberg service, it is possible to estimate a beta over many differing time periods, market indices, and as smoothed or unadjusted. The figures presented here represent the base-line or default-estimation approach used if other approaches are not specified.

Calculated Betas Change a lot

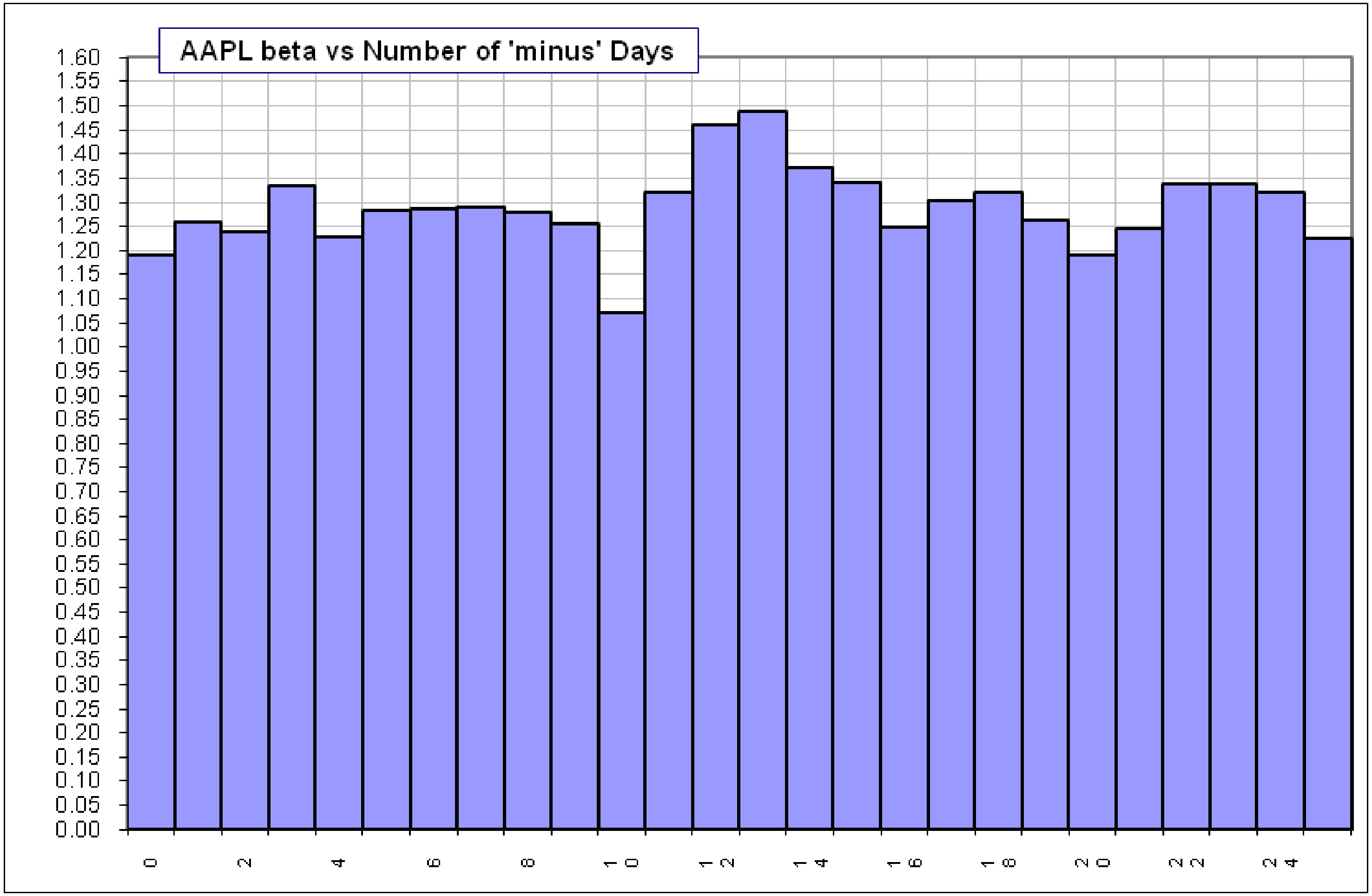
- Gummy and independently Fernandaz had an interesting idea.
- Suppose we are interested in calculating betas on monthly returns.
- The standard is to calculate it based on end of month data (last day of month).
- What if we use the 12th of each month, or 22nd, or something else. Will beta change?

Betas change from 1 day to the next

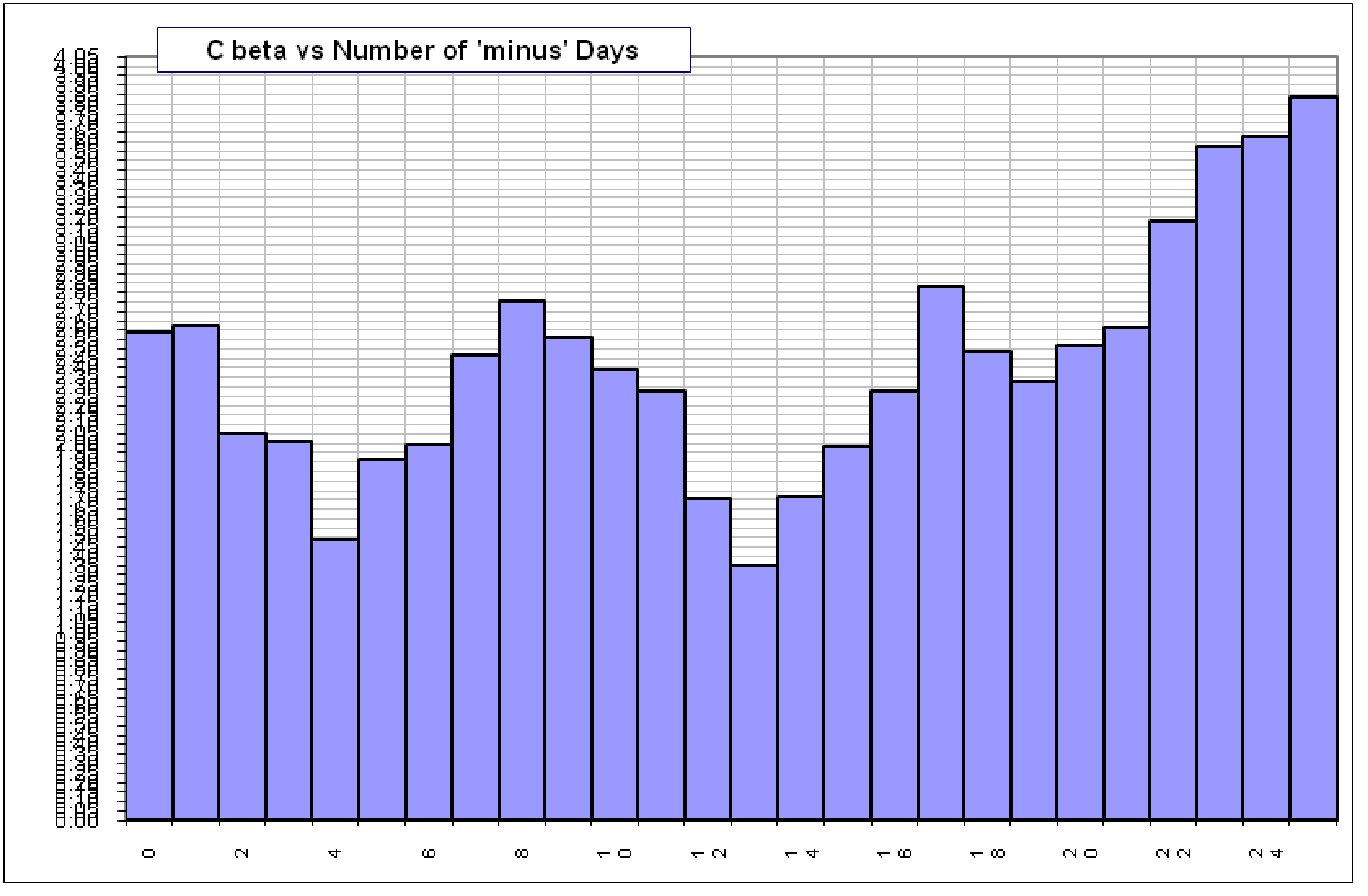
- Gummy has a great spreadsheet that calculates this **weird** result



AAPL Doesn't Change Too Much



But Check out Citibank



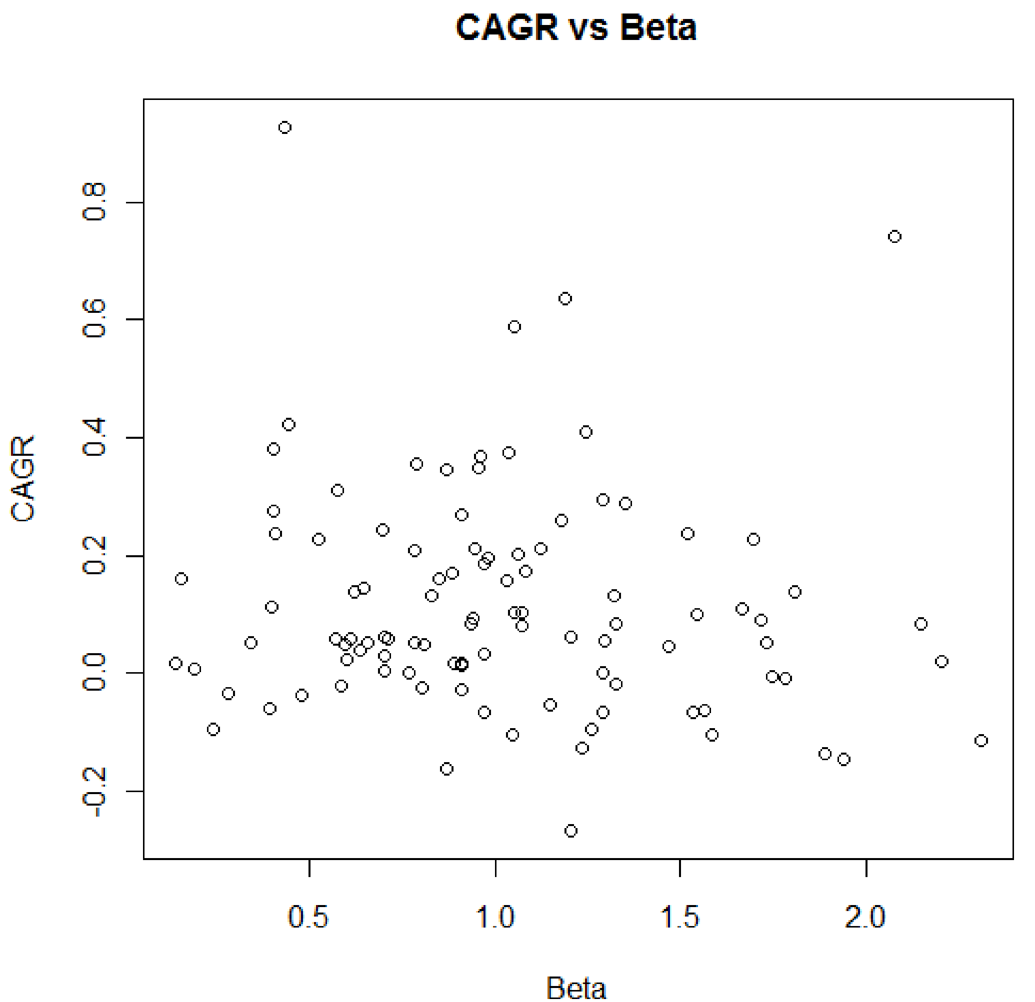
Conclusion

- Calculated Betas are very unstable!

CAGR and Beta

- According to the CAPM, a bigger Beta implies a bigger expected stock return.
- So what should a graph of CAGR (compounded annual return) versus Beta look like?
- Probably not what you expect.

Beta vs CAGR for Nasdaq 100 stocks



To be fair, we should see what this looks like during a bull market like 1996-2000, but it will look even worse.

Conclusion

- Calculated Betas are very unstable!
- So can't really trust the market model
- What is one to do? Read the appendix in "Are calculated betas worth for anything?"; it gives a very thorough analysis of the research on Beta through the years.

Adjusted Betas

- Betas are non-stationary.
- Researchers have shown that adjusted beta is often a better forecast of future beta than is historical beta. As a consequence, practitioners often use adjusted beta.

Beta as an Autoregressive Process

- One simple idea is to model beta as a first order autoregressive process:

$$\beta_{i,t+1} = \alpha_0 + \alpha_1 \beta_{i,t} + \epsilon_{i,t+1}$$

- It has been academically shown that adjusted beats predict future betas better than historical data since betas are on average mean reverting.

A very simple adjustment

- An adjustment that Merril Lynch and Bloomberg use is as follows.
- Adjusted Beta = 0.333+0.667*hist Beta
- Historical beta=1, adj beta =1
- Historical beta=1.5, adj beta = 1.33
- Historical beta=0.5, adj beta=0.667

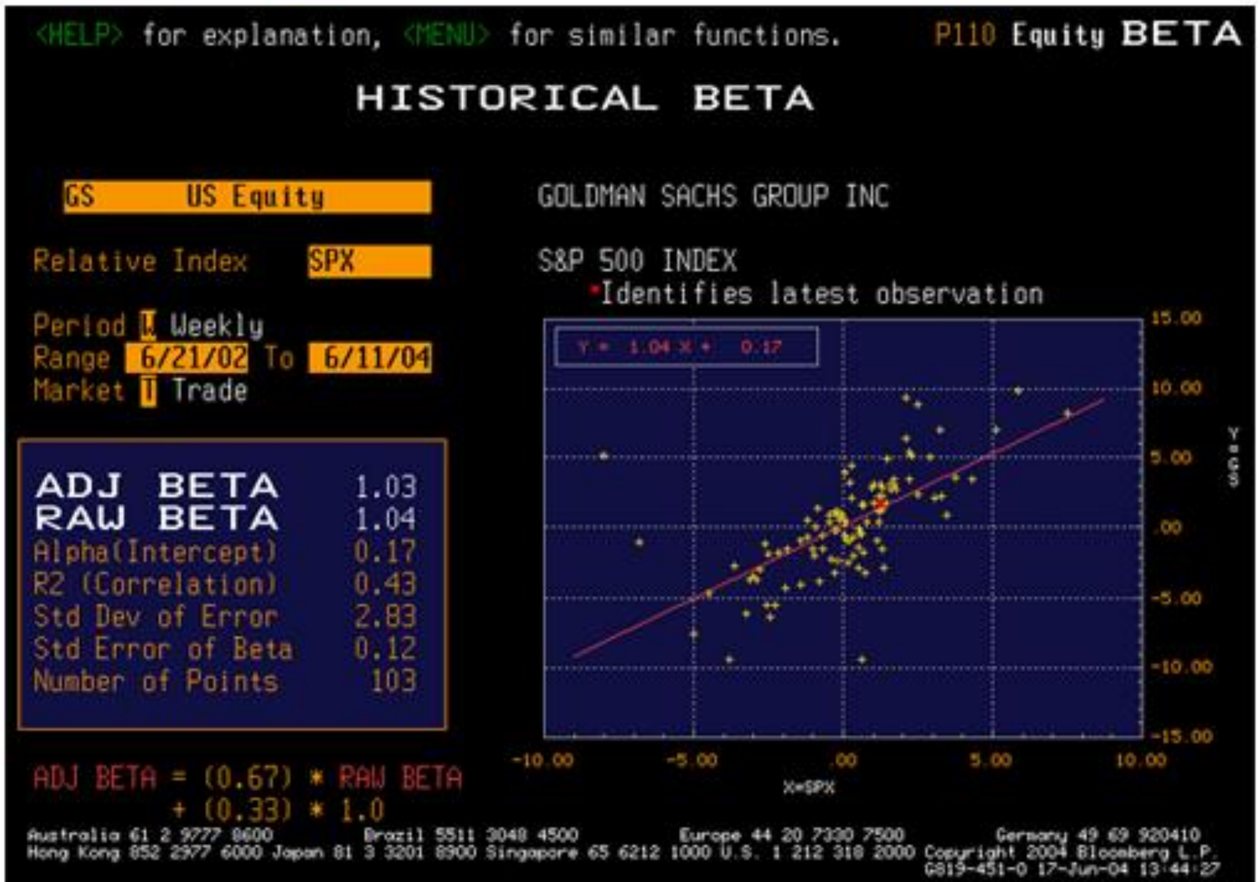
From Bloomberg

Enter this command to look up the beta of a stock:

{ticker symbol} EQUITY BETA GO

Example: The following screenshot shows the result for the beta of Goldman Sachs.

GS EQUITY BETA GO



Some Silliness

- Bloomberg adj beta = 0.333+0.667*beta
- Value Line adj beta = 0.35 + 0.67*beta
- Merril Lynch adj beta = 0.33743+0.66257*beta
- And there are many more methods out there…
- It is highly believed though that some adjustment of beta should be used instead of using the historical value directly.
- These adjusted betas are then used as before to estimate variances and covariances.

The Single Factor (or Index) Model

- The single factor model assumes that asset returns are correlated for only one reason.
- Each asset is assumed to respond (some in large fashion, some in small fashion) to the pull of a single factor, which is usually taken to be the market portfolio.

The Single Factor (or Index) Model

- The Single Index Model says

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

- where

α_i = part of the security's return independent of the market

R_M = rate of return on the market index

β_i = a constant that measures the expected change in R_i given a change in R_M

ε_i = random noise, indepent of R_M (firm-specific surprises)

- And one additional assumption…..

Discuss the model

- The single factor model implicitly assumes that two types of events produce the period-to-period variability in a stocks rate or return.
- The first type of event is called a **macro** event.
- Examples might include an unexpected change in the rate of inflation, a Fed rate cut, a war, etc….

Macro Events

- Macro events are broad or sweeping in their impact.
- They affect nearly all firms to one degree or another, and they may have an affect on the general level of stock prices.
- They produce a change in the rate of return of the market, and through the pull of the market, induce changes in the rates of returns on individual securities.

Account for Macro events individually?

- Alternatively, one may want to account for the macro factors individually.
- That is, separate factors for inflation, Fed rate, world events, etc…
- In this case one would use a **multifactor** model to be discussed later.

Micro Events

- The second type of event that produces uncertainty in a security's return in the single factor model is micro in nature.
- Micro events have an impact on individual firms but no generalized impact on other firms.
- Examples might include the discovery of a new product, a local labor strike, or resignation of a key person in the firm.

Micro Events

- These events are assumed to have no effect on other firms, and they have no effect on the value of the market portfolio or its rate of return.
- Micro events do affect the rate of return of the individual security, however.

Assumed Away

- Industry events have been assumed away by the single factor model.
- This is an event that has a generalized impact on many of the firms in a given industry but is not broad or important enough to have a significant impact on the general economy or the value of the market portfolio.
- This assumption is expressed mathematically on the next slide.

A Key Assumption

- A key assumption of this model is that for two different securities *i* and *j*, the noise components are unrelated.
- That is

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$$

Firm specific surprises for company i

Firm specific surprises for company j

What does this imply?

- This implies that the only reason stocks vary together, systematically, is because of a common co-movement with the market.
- That is, there are no effects beyond the market (e.g. industry effects) that account for co-movement of securities.
- This assumption is relaxed in more involved multi-index models.

Is the assumption true?

- It is obvious that the lack of within industry correlation isn't strictly accurate.
- After all, suppose something good happens to AAPL.
- This has an immediate impact on only on AAPL but also on the company's suppliers and competitors.
- Many companies would be affected simultaneously, some positively and other negatively.

Is the assumption true?

- The assumption says there are no industry effects between two securities.
- The residuals for these firms would not be independent but rater would be generated by some common event.
- We know therefore that the residuals are correlated to some degree.
- Its hoped however that the degree of correlation is small enough that the inaccuracy of the single-factor model’s portfolio variance equation doesn’t transcend is relative efficiency.
- One we can easily check empirically this assumption, also.

AAPL and MSFT

```
> getSymbols("AAPL",from="2008-03-01")
[1] "AAPL"
> getSymbols("SPY",from="2008-03-01")
[1] "SPY"
> spyret=monthlyReturn(SPY)
> aaplret=monthlyReturn(AAPL)
>
> fitaapl = lm(aaplret~spyret)
> residaapl = residuals(fitaapl)
```

```
> cor(residaapl,residmsft)
[1] -0.06842474
```

```
> getSymbols("MSFT",from="2008-03-01")
[1] "MSFT"
> getSymbols("SPY",from="2008-03-01")
[1] "SPY"
> spyret=monthlyReturn(SPY)
> msftret=monthlyReturn(MSFT)
>
> fitmsft = lm(msftret~spyret)
> residmsft = residuals(fitmsft)
```

MSFT and SBUX

- Same idea, do the market model regression for each stock, and see if the residuals are correlated with each other:

```
> cor(residsbux,residmsft)
[1] -0.0265771
```

AAPL and RIMM

- This one should be interesting since they are somewhat in the same industry

```
> cor(residaapl,residrimm)
[1] 0.4204353
```

- So slightly correlated (as expected one supposes). But hopefully not too strongly correlated to affect the entire model.

What does this give you?

$$\text{Suppose } \operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$$

- The math is a bit algebraic:

$$Cov(R_i, R_j) = Cov(\alpha_i + \beta_i R_M + \varepsilon_i, \alpha_j + \beta_j R_M + \varepsilon_j)$$

$$= \text{a bunch of terms}$$

$$= \beta_i \beta_j Cov(R_M, R_M)$$

$$= \sigma_M^2 \beta_i \beta_j$$

Uh, you're losing me

- Under the Single Index Model assumptions for stock returns, we have the following results (*i* and *j* index different stocks):

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_{\varepsilon,i}^2$$

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

- So what, you're thinking!
- **Hang on and we'll explain why this is useful**.

Estimating Inputs for Portfolios

- Mean Variance Portfolio estimation seems to be a good way to create portfolios.
- However, there remains the issue of how to estimate the required inputs.
- In the following slides we review
  - Historical means, variance and correlations
  - Historical Betas
  - Adjusted Betas

Historical Estimates

- This approach involves calculating means, variances, and correlations directly from historical data.
- The historical method requires estimating a very large number of parameters when we are optimizing for even a moderately small number of assets.

Historical Estimates (cont)

- The number of parameters a portfolio manager needs to estimate to determine the minimum-variance frontier depends on the number of potential stocks in the portfolio.
- If a portfolio manager has n stocks in a portfolio and wants to use mean–variance analysis, she must estimate
  - n parameters for the expected returns to the stocks,
  - n parameters for the variances of the stock returns, and
  - $n(n-1)/2$ parameters for the covariances of all the stock returns with each other.

  Together, the parameters total $n^2/2 + 3n/2$.

The Historical Approach

- The two limitations of the historical approach involve the quantity of estimates needed and the quality of historical estimates of inputs.
- The quantity of estimates needed may easily be very large, mainly because the number of covariances increases in the square of the number of securities.
- If the portfolio manager wanted to compute the minimum-variance frontier for a portfolio of 100 stocks, she would need to estimate 5, 150 parameters. If she wanted to compute the minimum-variance frontier for 1,000 stocks, she would need to estimate 501,500 parameters.
- Not only is this task unappealing, it might be impossible. (the number of time series observations needs to exceed the number of assets we have).
- The second limitation is that historical estimates of return parameters typically have substantial estimation error

Single Index Model Estimates

- Under the Single Index Model assumptions for stock returns, we have the following results:

$$E(R_i) = \alpha_i + \beta_i E(R_M)$$

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_{\varepsilon,i}^2$$

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

What does this get us?

- We can use the single index model to greatly reduce the computational task of providing the inputs to a mean–variance optimization.
- For each of the *n* assets, we need to know

$$\alpha_i, \beta_i, \sigma_{\varepsilon,i}^2, \mu_M, \sigma_M^2$$

- Thus we only need to estimate 3*n*+2 parameters using the single index model.

What does this get us?

- We need far fewer parameters to construct the minimum-variance frontier than we would if we estimated the historical means, variances, and covariances of asset returns.
- For example, if we estimated the minimum-variance frontier for 1,000 assets (say, 1,000 different stocks), the market model would use 3,002 parameters for computing the minimum-variance frontier, whereas the historical estimates approach would require 501,500 parameters, as discussed earlier.

Simple Example

- Lets compute the covariance between AAPL and SBUX using the single index model, and compare it to the empirical estimate.
- We will need to
  - Collect data
  - Run two market model regressions
  - Estimate

Review the two equations

- The two relevant equations are

$$\mathrm{Var}(R_i) = \beta_i^2 \sigma_M^2 + \sigma_{\epsilon_i}^2$$

Variance of residual noise estimated from regression model

Slope from regression model

Variance of market estimated directly

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

Obtain the Historical Data

- Using R

```
getSymbols("AAPL")
getSymbols("SPY")
getSymbols("SBUX")
sbuxret=monthlyReturn(SBUX)
aaplret=monthlyReturn(AAPL)
spyret=monthlyReturn(SPY)
```

Obtain σ_M^2

- Obtain the variance of market returns

```
spyvar=var(spyret)
```

Obtain the betas

- Run the market model and obtain beta

```
fitaapl=lm(aaplret~spyret)
fitsbux=lm(sbuxret~spyret)

aaplbeta=fitaapl$coef[2]

sbuxbeta=fitsbux$coef[2]
```

Obtain $\sigma^2_{\epsilon_i}$

- Obtain the standard deviation of residual noise

```
s_aapl = summary(fitaapl)$sigma
s_sbux = summary(fitsbux)$sigma
```

- We would square these of course to get the variance.

Time to Estimate $\mathrm{Var}(R_i) = \beta_i^2 \sigma_M^2 + \sigma_{\epsilon_i}^2$

- Using the formula

$$\mathrm{Var}(R_i) = \beta_i^2 \sigma_M^2 + \sigma_{\epsilon_i}^2$$

```
varaapl = (aaplbeta^2)*(spyvar)+(s_aapl^2)

varsbux = (sbuxbeta^2)*(spyvar)+(s_sbux^2)
```

Compare

- How do the market model variance estimates compare with the empirical estimates?

```
> var(aaplret)
                monthly.returns
monthly.returns      0.01140558
> varaapl
                monthly.returns
monthly.returns      0.01151711

> var(sbuxret)
                monthly.returns
monthly.returns      0.00980832
> varsbux
                monthly.returns
monthly.returns     0.009901813
```

Compute the covariance

- Using the formula

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

```
cov_aapl_sbux = aaplbeta*sbuxbeta*spyvar
```

Compare

- Compare the empirical covariance versus the one calculated using the market model:

```
> cov(aaplret,sbuxret)
                monthly.returns
monthly.returns     0.004881556
> cov_aapl_sbux
                monthly.returns
monthly.returns     0.004346767
```

I’m still lost

- Suppose we are back at the efficient portfolio problem, and want to create the required inputs to find the efficient frontier.
- Recall that we need the mean returns for all assets, and the variance-covariance matrix.
- If we have N assets, we need to compute N(N-1)/2 covariances.

Covariance Estimation is Hard

- There are many elements to estimate in a covariance/correlation matrix.
- Introducing a model that simplifies the estimation process is very helpful.
- In addition, if the model can use system-wide (market-wide) versus firm-specific sources of variation, the problem becomes much more manageable.

A little less lost

- The covariance estimation problem has become a little easier.
- We simply run N market model regressions and use the following formulas

$$\hat{\sigma}_i^2 = \hat{\beta}_i^2 \hat{\sigma}_M^2 + \hat{\sigma}_{\varepsilon,i}^2$$

$$[Cov(R_i, R_j)] = \hat{\sigma}_{ij} = \hat{\sigma}_M^2 \hat{\beta}_i \hat{\beta}_j$$

- These formulas require NO knowledge of the co-movement of any two stocks, just how they each move with the market

Example

- A four stock example of AAPL, JNJ, SBUX and MSFT.
- We are going to compare the empirical variance-covariance matrix estimated directly from the data, with the one created with the Simple Index Model.
- This is a very basic example; in practice one might do this with N=250 stocks.

The Code

```
getSymbols("AAPL",from="2008-03-01")
getSymbols("JNJ",from="2008-03-01")
getSymbols("SBUX",from="2008-03-01")
getSymbols("MSFT",from="2008-03-01")
getSymbols("SPY",from="2008-03-01")

spyret=monthlyReturn(Ad(SPY))
aaplret=monthlyReturn(Ad(AAPL))
jnjret=monthlyReturn(Ad(JNJ))
msftret=monthlyReturn(Ad(MSFT))
sbuxret=monthlyReturn(Ad(SBUX))

aaplbeta=coef(lm(aaplret~spyret))[2]
jnjbeta=coef(lm(jnjret~spyret))[2]
msftbeta=coef(lm(msftret~spyret))[2]
sbuxbeta=coef(lm(sbuxret~spyret))[2]

se_aapl=summary(lm(aaplret~spyret))$sigma
se_jnj=summary(lm(jnjret~spyret))$sigma
se_msft=summary(lm(msftret~spyret))$sigma
se_sbux=summary(lm(sbuxret~spyret))$sigma
```

The Code

```
se_aapl=summary(lm(aaplret~spyret))$sigma
se_jnj=summary(lm(jnjret~spyret))$sigma
se_msft=summary(lm(msftret~spyret))$sigma
se_sbux=summary(lm(sbuxret~spyret))$sigma


sbetas=c(aaplbeta,jnjbeta,msftbeta,sbuxbeta)
regvars = c(se_aapl^2,se_jnj^2,se_msft^2,se_sbux^2)
covmat=matrix(nrow=4,ncol=4)
rownames(covmat) = c("AAPL","JNJ","MSFT","SBUX")
colnames(covmat) = c("AAPL","JNJ","MSFT","SBUX")
varmarket = var(spyret)
for(i in 1:4)
  for(j in 1:4)
  covmat[i,j]=sbetas[i]*sbetas[j]*varmarket
for(i in 1:4)
  covmat[i,i] = covmat[i,i]+regvars[i]
covmat=round(covmat,6)

empcovmat=cov(cbind(aaplret,jnjret,msftret,sbuxret))
rownames(empcovmat) = c("AAPL","JNJ","MSFT","SBUX")
colnames(empcovmat) = c("AAPL","JNJ","MSFT","SBUX")
empcovmat=round(empcovmat,6)
```

How does it compare?

- Here is the covariance matrix computed from the historical returns

```
> empcovmat
         AAPL      JNJ      MSFT     SBUX
AAPL 0.007626 0.001069 0.002479 0.003526
JNJ  0.001069 0.001818 0.001221 0.001080
MSFT 0.002479 0.001221 0.004995 0.002329
SBUX 0.003526 0.001080 0.002329 0.007206
```

- And using Betas

```
> covmat
         AAPL      JNJ      MSFT     SBUX
AAPL 0.007682 0.001470 0.002333 0.002725
JNJ  0.001470 0.001829 0.001324 0.001547
MSFT 0.002333 0.001324 0.005027 0.002455
SBUX 0.002725 0.001547 0.002455 0.007254
```

Another Covariance Example

- A good question, “what if we had included two stocks in the same sector, how would the covariance matrix from the single index model differ from the empirical covariance matrix?”
- Let’s see what happens.
- The idea is that this might violate the assumption that

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$$

Another Covariance Example

- A four stock example of AAPL, HPQ, SBUX and MSFT.
- We are going to compare the empirical variance-covariance matrix estimated directly from the data, with the one created with the Simple Index Model.

HPQ and AAPL

Apple Inc. (AAPL) - NasdaqGS ★ **Watchlist**

119.50 ↓1.03 (0.85%) Oct 30, 4:00PM EDT

After Hours : **119.28** ↓0.22 (0.18%) Oct 30, 7:59PM EDT

Competitors

Get Competitors fo

Direct Competitor Comparison

	AAPL	HPQ	GOOG	PVT1	Industry
Market Cap:	666.25B	48.48B	488.84B	N/A	1.04B
Employees:	110,000	302,000	59,976	6,225[1]	980.00
Qtrly Rev Growth (yoy):	0.22	-0.08	0.13	N/A	0.16
Revenue (ttm):	233.72B	106.05B	71.76B	3.34B[1]	560.74M
Gross Margin (ttm):	0.40	0.24	0.63	N/A	0.21
EBITDA (ttm):	82.49B	12.76B	23.30B	N/A	25.06M
Operating Margin (ttm):	0.30	0.08	0.26	N/A	0.04
Net Income (ttm):	53.39B	4.56B	15.44B	-304.00M[1]	N/A
EPS (ttm):	9.22	2.44	23.63	N/A	0.05
P/E (ttm):	12.96	11.04	30.08	N/A	18.72
PEG (5 yr expected):	0.79	10.64	1.47	N/A	1.27
P/S (ttm):	2.88	0.46	6.87	N/A	1.35

The Resultant Covariance Matrices

- The Empirical Matrix

	AAPL	HPQ	MSFT	SBUX
AAPL	0.007626	0.001830	0.002479	0.003526
HPQ	0.001830	0.008524	0.003062	0.003014
MSFT	0.002479	0.003062	0.004995	0.002329
SBUX	0.003526	0.003014	0.002329	0.007206

- From the Single Index Model

	AAPL	HPQ	MSFT	SBUX
AAPL	0.007682	0.002847	0.002333	0.002725
HPQ	0.002847	0.008584	0.002566	0.002997
MSFT	0.002333	0.002566	0.005027	0.002455
SBUX	0.002725	0.002997	0.002455	0.007254

So possibly a bigger estimation error going on.