Stat 107: Introduction to Business and Financial Statistics

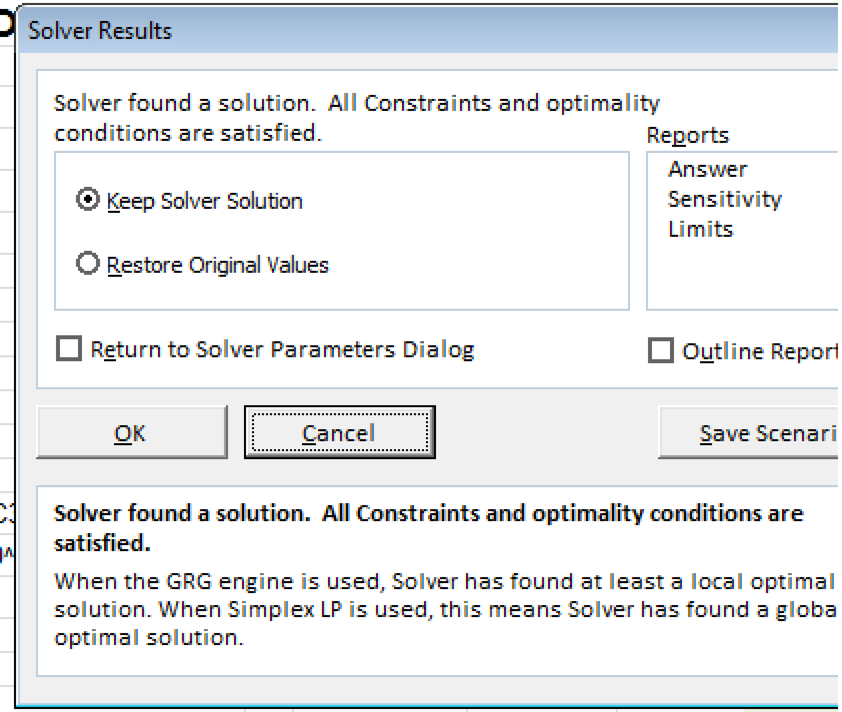
Class 11: Portfolios, Part IV

Some R functions

- portfolio.txt code on course web site
- Will compute
  - Efficient portfolio for a target return
  - The efficient frontier
  - Global Minimum Portfolio
  - Tangency Portfolio (discussed today)

Recall 2 Asset Example

	A	B	C	D
1	**LOCATING THE MINIMUM VARIANCE PORTFO**			
2		**K**	**XOM**	
3	Average, $E(r_K)$ and $E(r_{XOM})$	6.00	11.71	
4	Variance, $Var(r_K)$ and $Var(r_{XOM})$	1.71	2.67	
5	Sigma, σ_K and σ_{XOM}	1.31	1.63	
6	Covariance of returns, $Cov(r_K, r_{XOM})$	0.74		
7				
8	**Portfolio return and risk**			
9	Percentage in K	0.6655		
10	Percentage in XOM	0.3345		
11				
12	Constraint	1.0000	<-- =B9+B10	
13				
14	Expected portfolio return, $E(r_p)$	7.9099	<-- =B9*B3+B10*C3	
15	Portfolio variance, $Var(r_p)$	1.3856	<-- =B9^2*B4+B10^	
16	Portfolio standard deviation, σ_p	1.1771	<-- =SQRT(B15)	
17				
18				
19				



R Code for Global Minimum

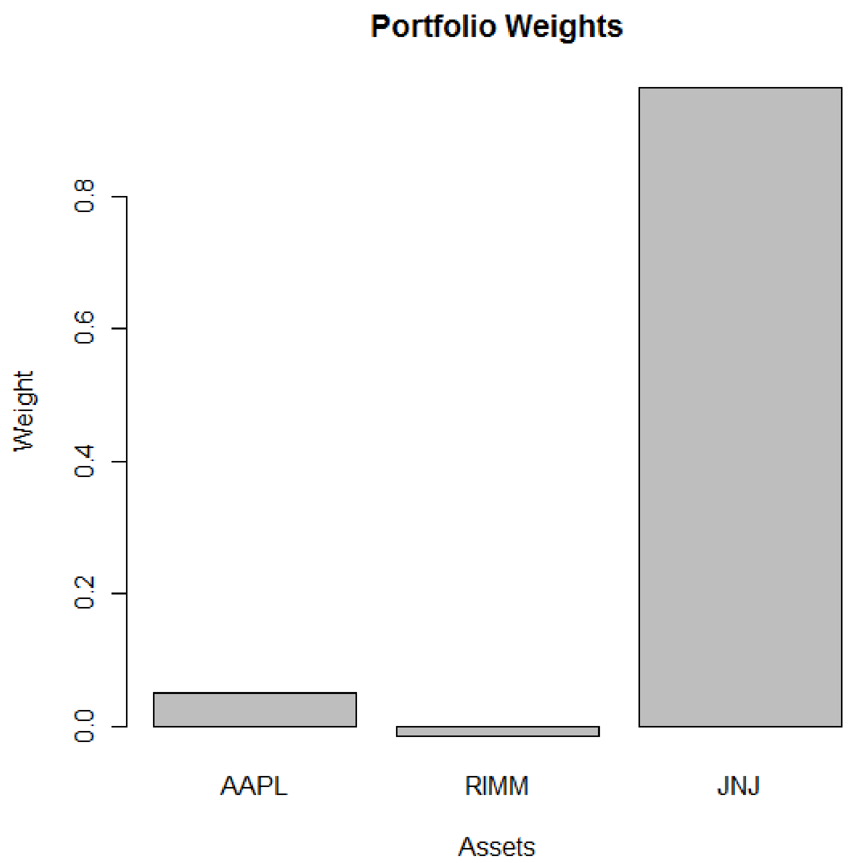
```
er=c(6,11.71)
covmat=matrix(nrow=2,ncol=2,c(1.71,.74,.74,2.67))
gmin.port <- globalMin.portfolio(er, covmat)
attributes(gmin.port)
print(gmin.port)
##plot(gmin.port)

> print(gmin.port)
Call:
globalMin.portfolio(er = er, cov.mat = covmat)

Portfolio expected return:     7.909897
Portfolio standard deviation:  1.177095
Portfolio weights:
[1] 0.6655 0.3345
```

R output for GMV (3 stocks)

```
> er
    AAPL     RIMM      JNJ
 0.03469 -0.01273  0.00071
> covmat
        AAPL    RIMM     JNJ
AAPL 0.01125 0.01096 0.00161
RIMM 0.01096 0.03904 0.00204
JNJ  0.00161 0.00204 0.00199
> gmin.port=globalMin.portfolio(er,covmat)
> print(gmin.port)
Call:
globalMin.portfolio(er = er, cov.mat = covmat)

Portfolio expected return:     0.002638269
Portfolio standard deviation:  0.04438313
Portfolio weights:
   AAPL    RIMM     JNJ
 0.0511 -0.0142  0.9631
```

R output for GMV

plot(gmin.port)

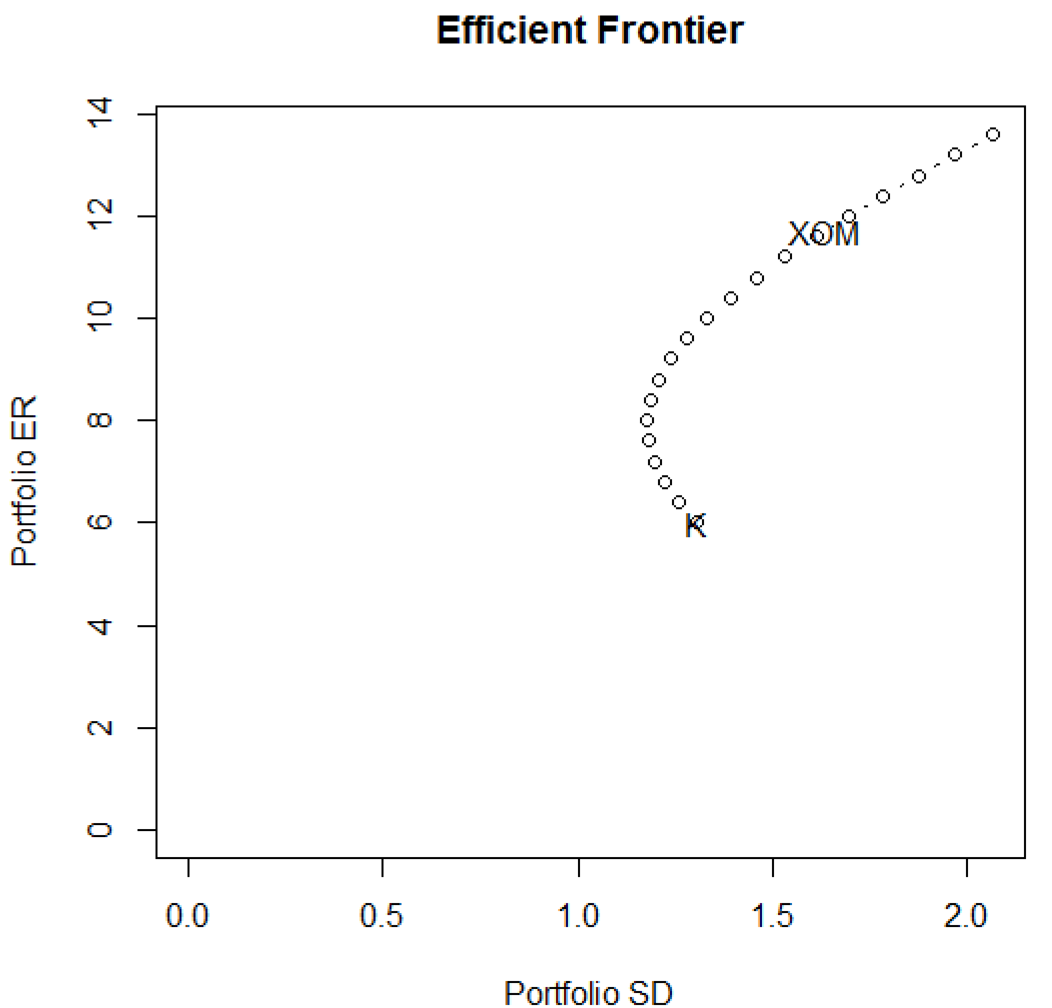


The efficient frontier in R

- The R code

```
> ef=efficient.frontier(er,covmat)
> plot(ef,plot.assets=T)
```

The 2 Stock Output



The 3 Stock Example Output

This image cannot currently be displayed.

The R Code for 3 Stocks

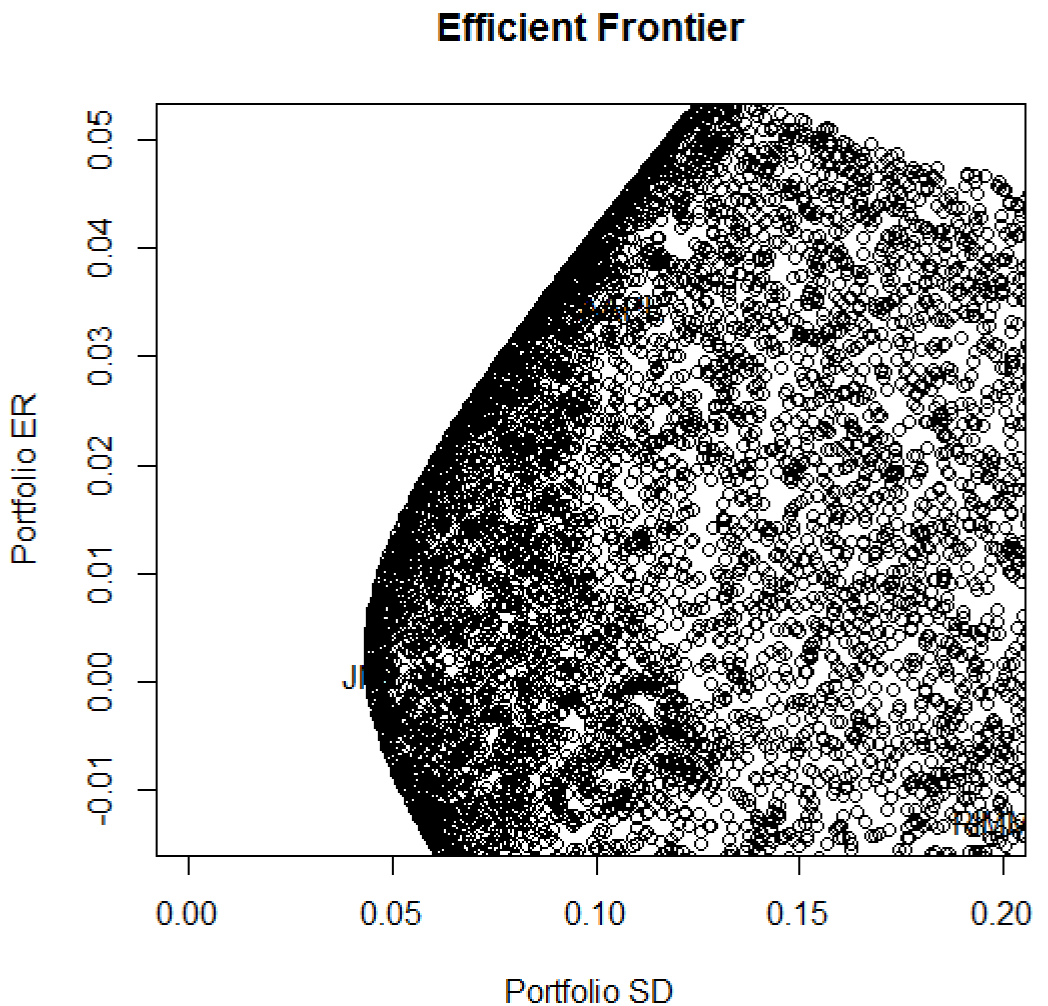
```
er=c(0.03469, -0.01273, 0.00071)
covmat=matrix(c(0.01125, 0.01096, 0.00161,0.01096,
0.03904, 0.00204,0.00161,
0.00204, 0.00199),nrow=3,ncol=3)
names(er)=c("AAPL","RIMM","JNJ")
colnames(covmat)=c("AAPL","RIMM","JNJ")
rownames(covmat)=c("AAPL","RIMM","JNJ")

gmin.port <- globalMin.portfolio(er, covmat)
attributes(gmin.port)
print(gmin.port)
summary(gmin.port, risk.free=rk.free)
plot(gmin.port)

ef=efficient.frontier(er,covmat)
plot(ef,plot.assets=T)
```

Can Also create random portfolios

- There is a getPortfolio function

```
for(i in 1:10000) {
w1=runif(1,min=-.5,1.5)
w2=runif(1,min=-.5,1.5)
foo=getPortfolio(er,covmat,c(w1,w2,1-w1-w2))
points(foo$sd,foo$er)
}
```

Resulting Graph



Can do this with many stocks

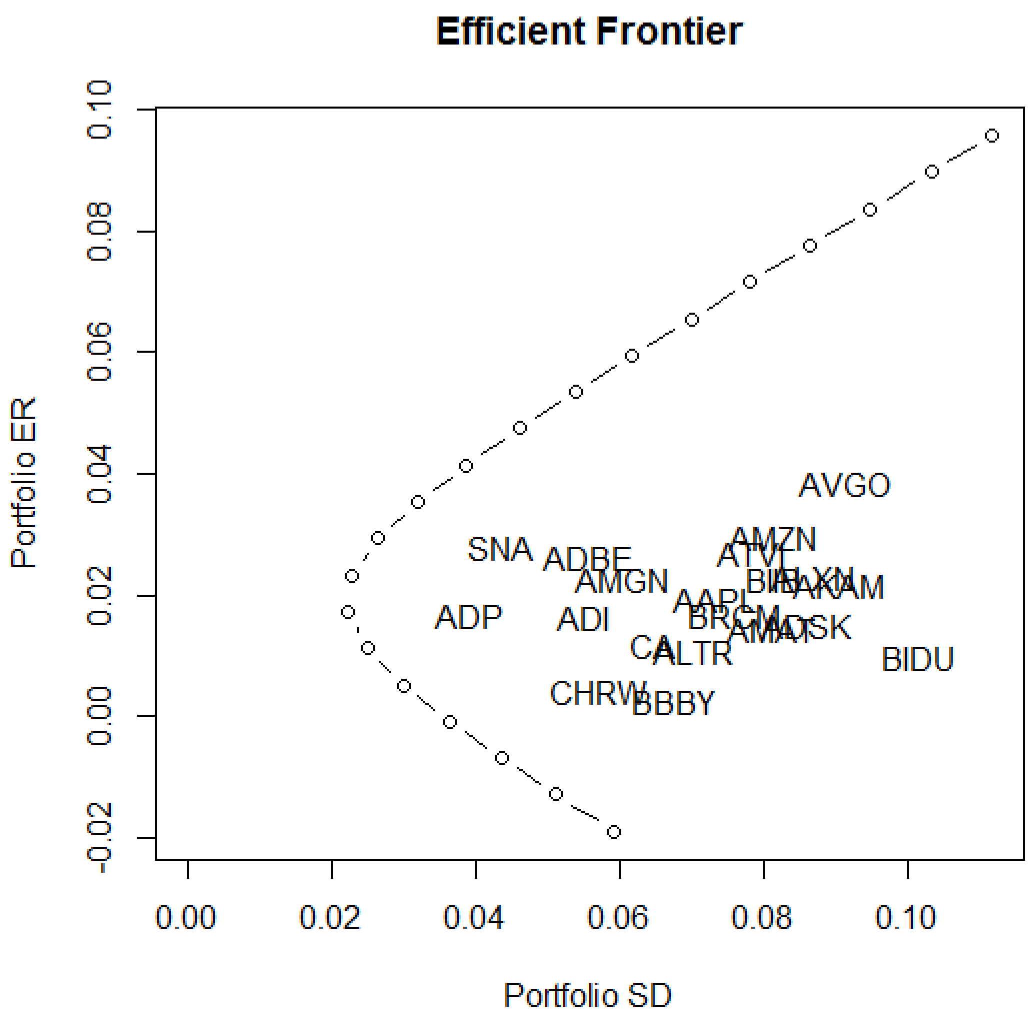
```
mystocks=read.csv("http://people.fas.harvard.edu/~mparzen/stat107/nasdaq1.csv",header
=FALSE, colClasses="character")
mystocks=mystocks[1:20,1]
n=length(mystocks)
getSymbols("SPY",from="2012-01-01")
nn=nrow(monthlyReturn(SPY))
myrets=matrix(nrow=nn,ncol=20)
for(i in 1:20) {
cat("I = ",i,"\n")
 x = getSymbols(mystocks[i],from="2012-01-01",auto.assign=FALSE)
 myrets[,i]=monthlyReturn(Ad(x))
}


er = apply(myrets,2,mean)
covmat = cov(myrets)
names(er)=mystocks

ef=efficient.frontier(er,covmat,alpha.min=-3,alpha.max=3)
plot(ef,plot.assets=T)
```

The GMV

- Easy for R to compute

```
> globalMin.portfolio(er, covmat)
Call:
globalMin.portfolio(er = er, cov.mat = covmat)

Portfolio expected return:     0.01931961
Portfolio standard deviation:  0.02204853
Portfolio weights:
   ATVI    ADBE    AKAM    ALXN    ALTR    AMZN    AMGN     ADI    AAPL    AMAT
 0.0294  0.1196 -0.0292 -0.0070  0.1254  0.0166  0.0064  0.0268 -0.0280 -0.2105
   ADSK     ADP    AVGO    BIDU    BBBY    BIIB    BRCM    CHRW      CA     SNA
-0.1527  0.2434 -0.0119 -0.0007 -0.0193  0.1514  0.0791  0.2589  0.0673  0.3353
```

The Efficient Fronteir

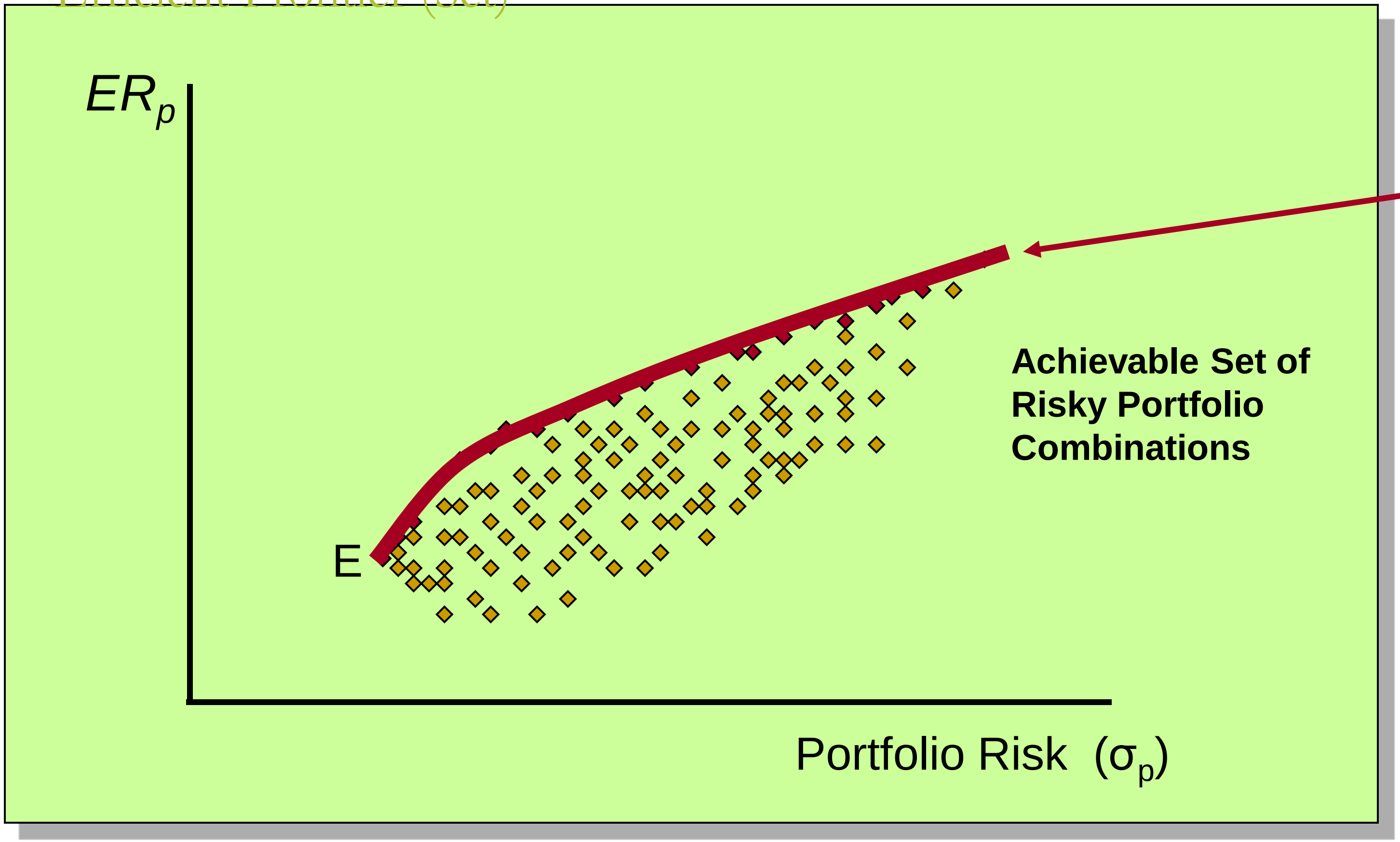


What now?

- The efficient frontier displays all the feasible portfolios possible, given any number of assets.
- However, **one can do even better** by introducing a risk-free asset into the mix.
- We will now describe the new, or super, efficient frontier.

Recall: Achievable Portfolio Combinations

Efficient Frontier (Set)



Efficient frontier is the set of achievable portfolio combinations that offer the highest rate of return for a given level of risk.

A Risk-Free Asset

- Now consider a risk-free asset
- Such an asset is generally consider a T-bill or Treasury Bill, or money in the bank.
- Let r_f denote the return of the risk-free asset
- Then $Var(r_f)$ =0 (its risk-free!).

Risk-free Investing

- When we introduce the presence of a risk-free investment, a whole new set of portfolio combinations becomes possible.
- We can estimate the return on a portfolio made up of RF asset (r_f) and a risky asset A letting the weight (*w*) invested in the risky asset and the weight invested in RF as (1-*w*)
- Our portfolio return: $R_p = (1-w)(r_f) + wR_A$

The Risk-Free Asset

- Note that by definition
  - $E[r_f] = r_f$
  - $Var[r_f] = 0$
  - $Cov(A, r_f) = 0$ for any other portfolio A

The New Efficient Frontier

Risk-Free Investing

- Expected return and risk on a two asset portfolio made up of risky asset *A* and *RF*:

$$R_p = (1\text{-}w)(r_f) + wR_A\text{, so}$$

$$E[R_p] = (1\text{-}w)(r_f) + wE[R_A]$$

$$\sigma_P^2 = w^2 \sigma_A^2$$

$$\sigma_P = w\sigma_A$$

Equation of a line

- From

$$E[R_p] = (1\text{-}w)(r_f) + wE[R_A]$$

$$\sigma_P = w\sigma_A$$

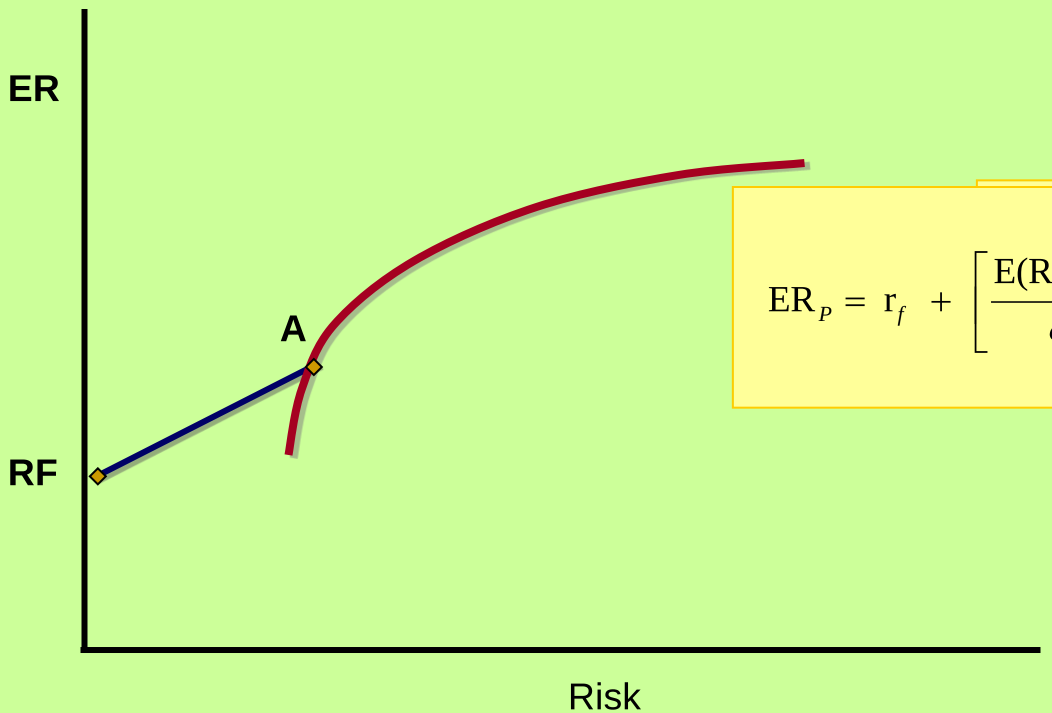
- We obtain

$$E[R_p] = r_f + w(E[R_A]\text{-}r_f)$$

$$= r_f + \sigma_p \left(\frac{E[R_A]\text{-}r_f}{\sigma_A} \right)$$

The New Efficient Frontier

Attainable Portfolios Using *RF* and *A*

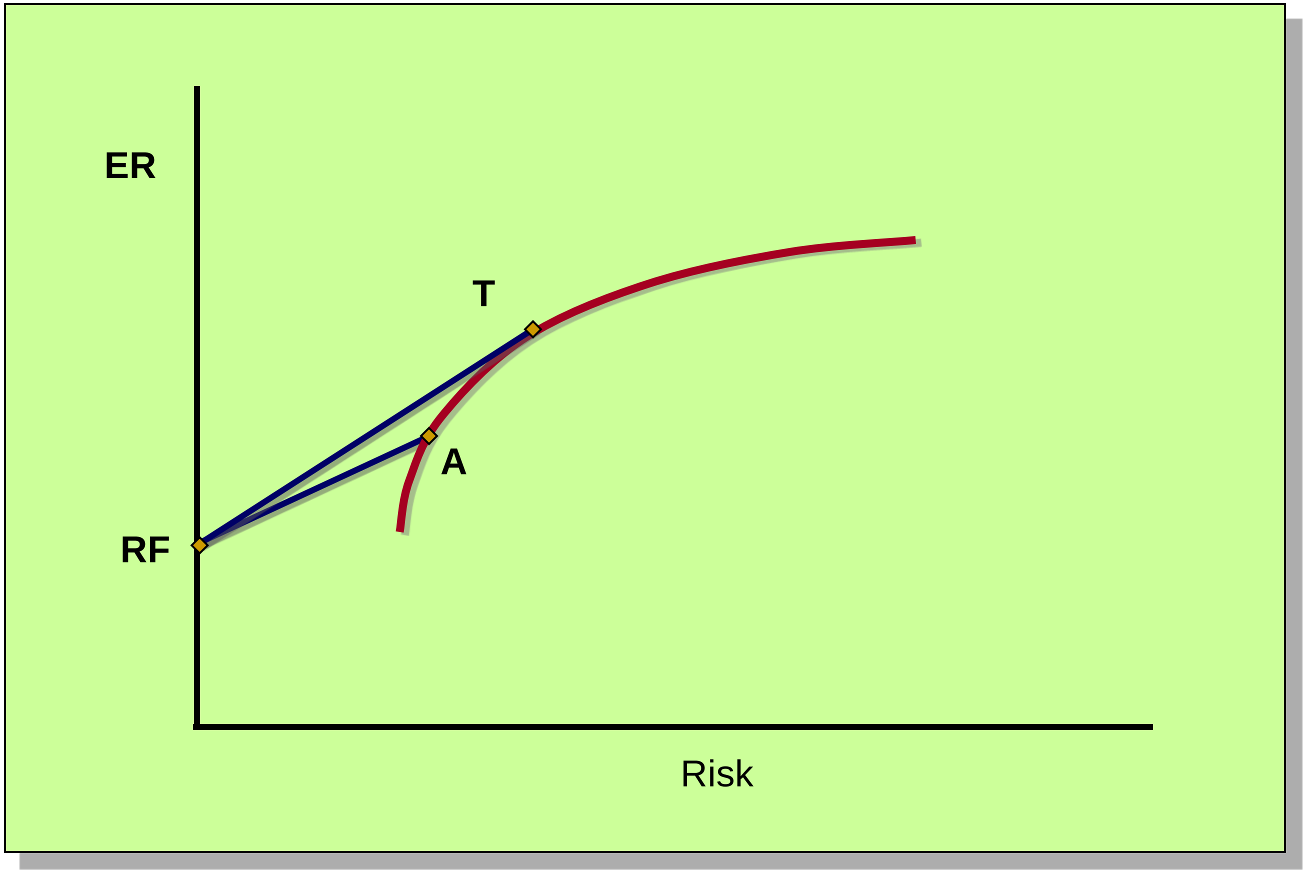


$$ER_P = r_f + \left[\frac{E(R_A) - r_f}{\sigma_A}\right]\sigma_P$$

This means you can achieve any portfolio combination along the blue coloured line simply by changing the relative weight of *RF* and *A* in the two asset portfolio.

The New Efficient Frontier

Attainable Portfolios using the *RF* and *A*, and *RF* and *T*



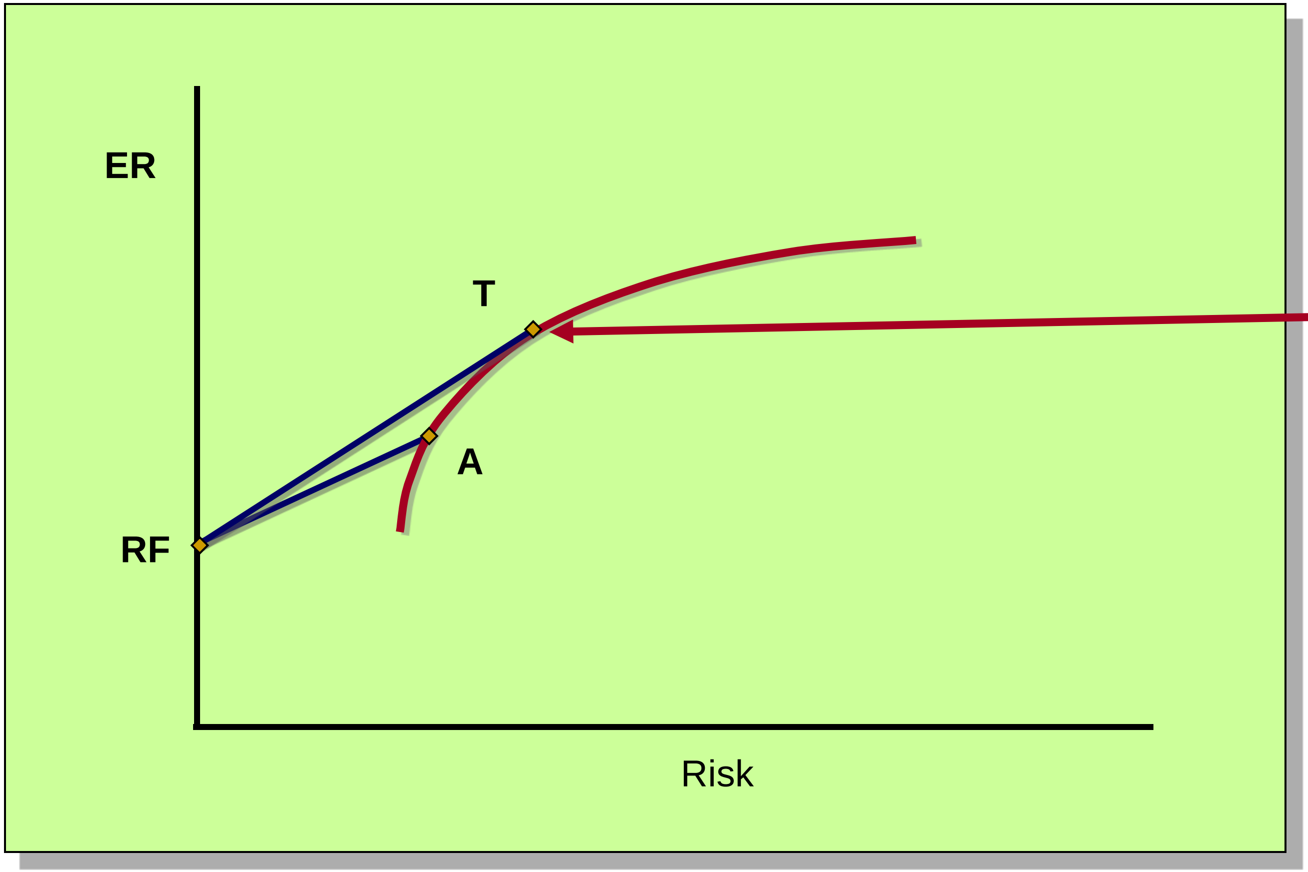
Which risky portfolio would a rational risk-averse investor choose in the presence of a *RF* investment?

Portfolio *A*?

Tangent Portfolio *T*?

The New Efficient Frontier

Efficient Portfolios using the Tangent Portfolio *T*

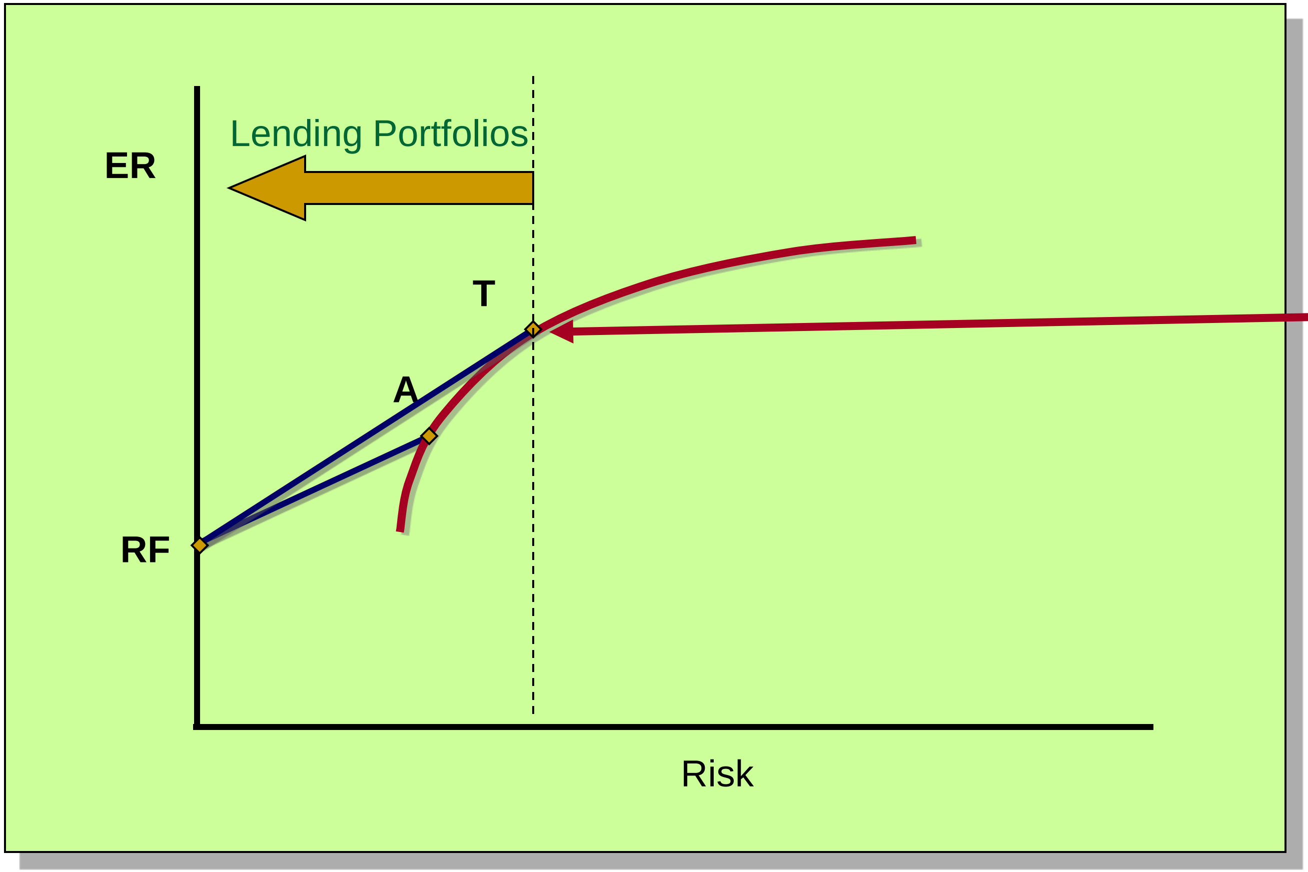


Clearly *RF* with *T* (the tangent portfolio) offers a series of portfolio combinations that dominate those produced by *RF* and *A*.

Further, they dominate all but one portfolio on the efficient frontier!

The New Efficient Frontier

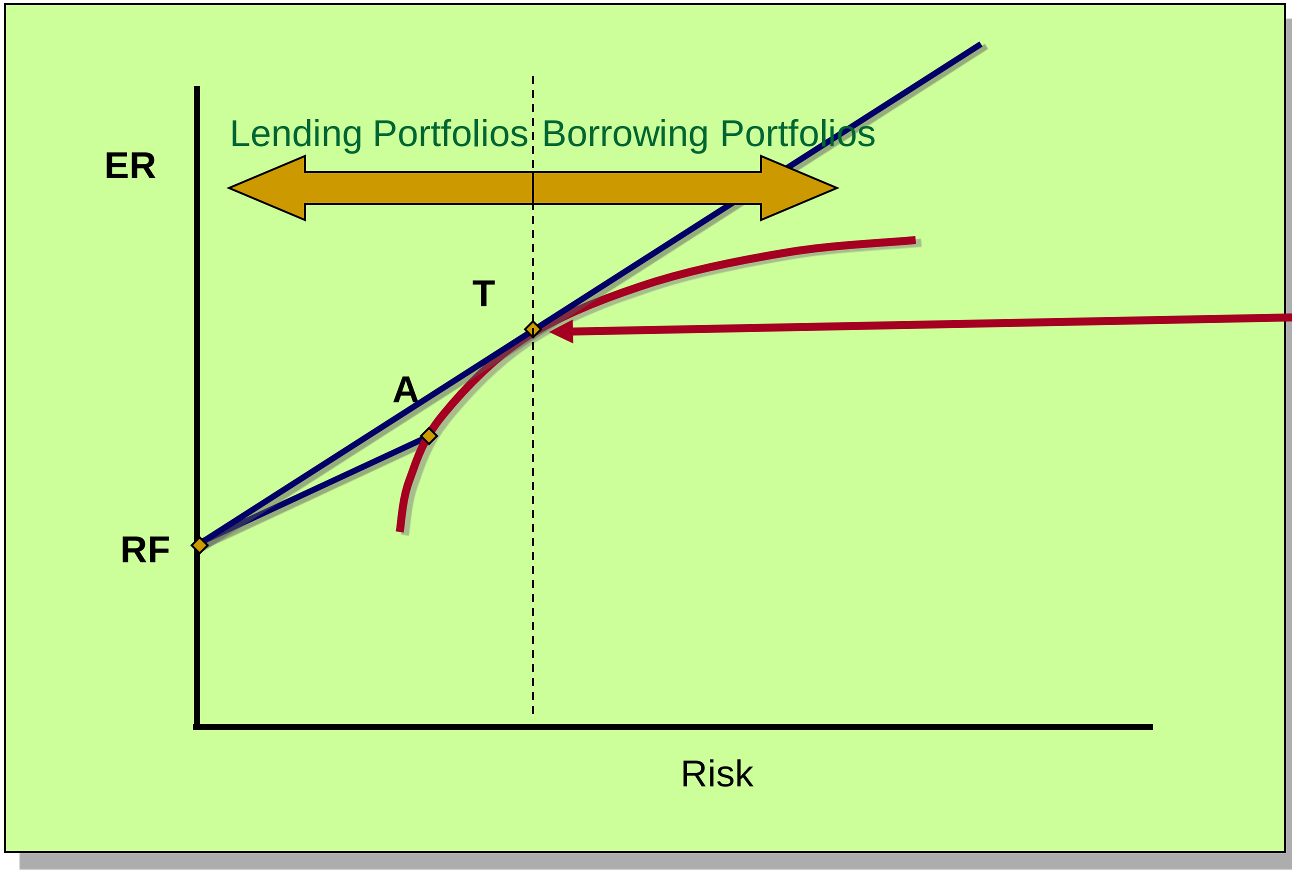
Lending Portfolios



Portfolios between *RF* and *T* are 'lending' portfolios, because they are achieved by investing in the Tangent Portfolio and lending funds to the government (purchasing a T-bill, the *RF*).

The New Efficient Frontier

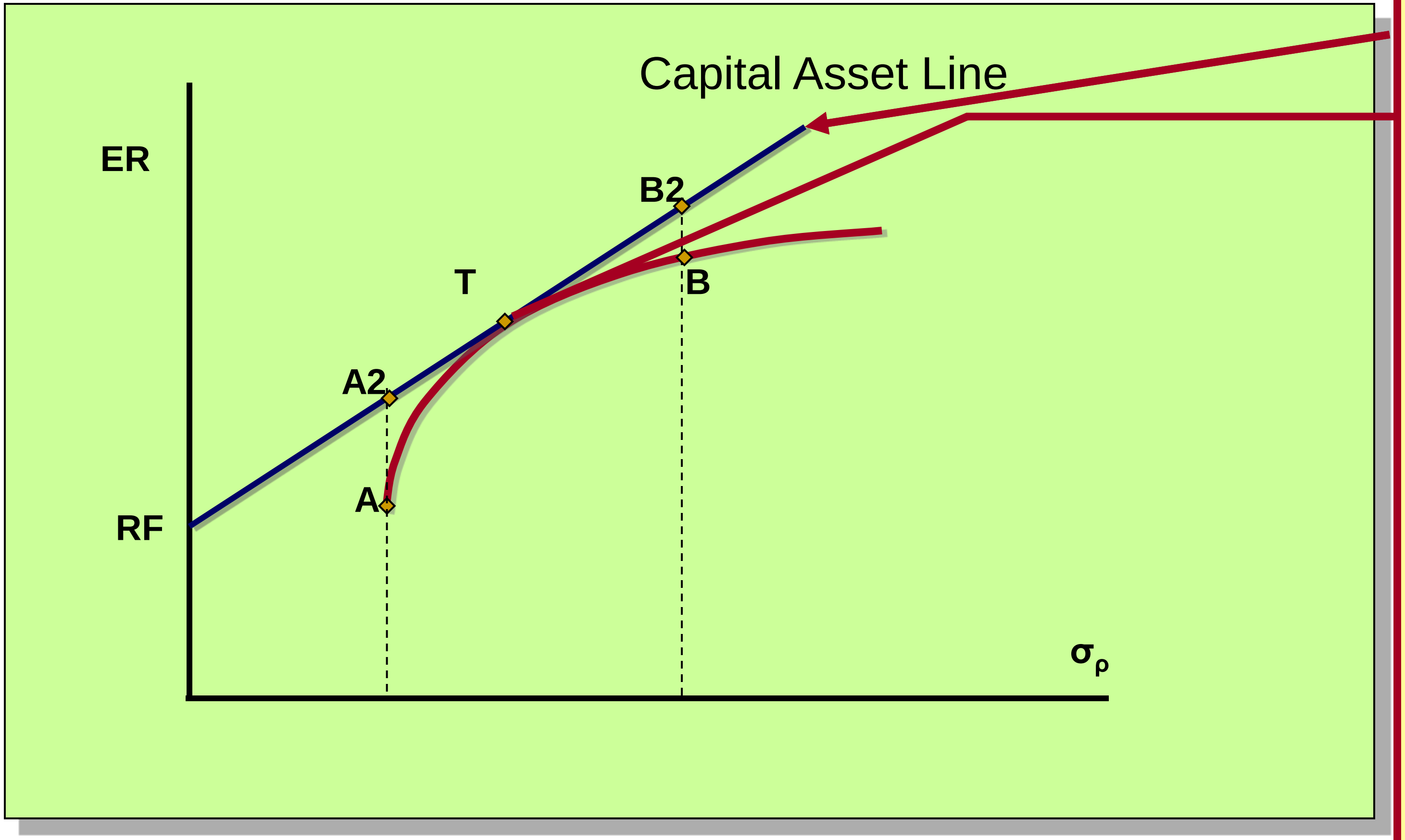
Borrowing Portfolios



The line can be extended to risk levels beyond 'T' by borrowing at RF and investing it in T. This is a levered investment that increases both risk and expected return of the portfolio.

The New Efficient Frontier

The New (Super) Efficient Frontier



This is now called the new (or super) efficient frontier of risky portfolios.

Investors can achieve any one of these portfolio combinations by borrowing or investing in RF in combination with the tangent portfolio.

The Market Portfolio

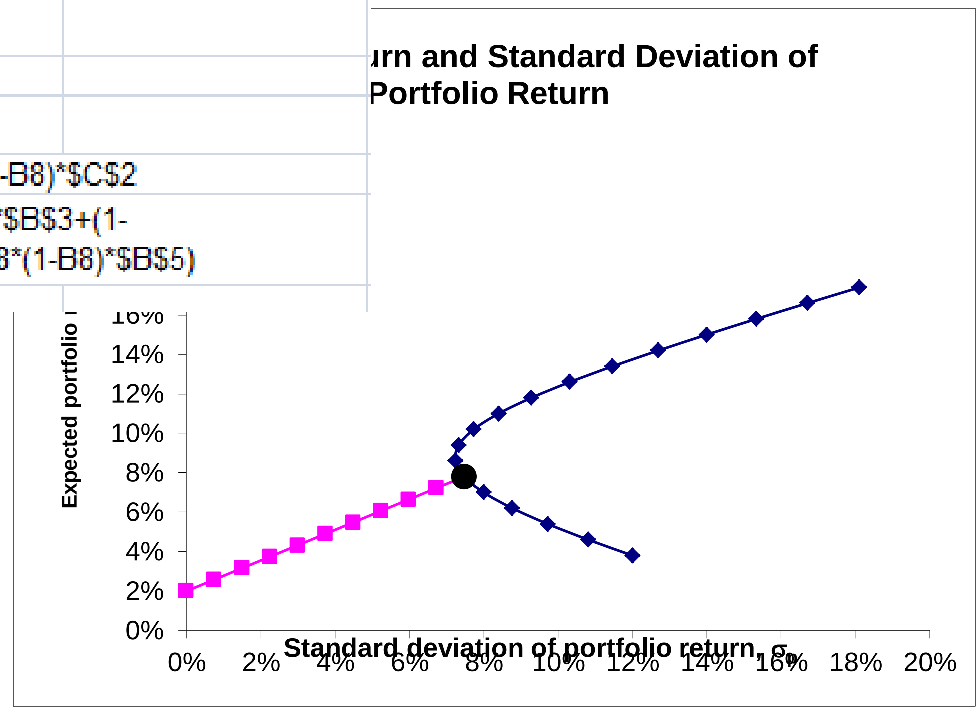
- Portfolio T, called the tangent portfolio, is also called the Market Portfolio
- The Capital Market Line is tangent to the efficient frontier, so every portfolio on the efficient frontier is below this line.
- ALL the best investment portfolios are on this line.

The Capital Market Line

- Is the set of optimal portfolio investments
- Each point on the line is
  - A combination of some percentage invested in the risk free asset
  - Another percentage invested in the market portfolio M.

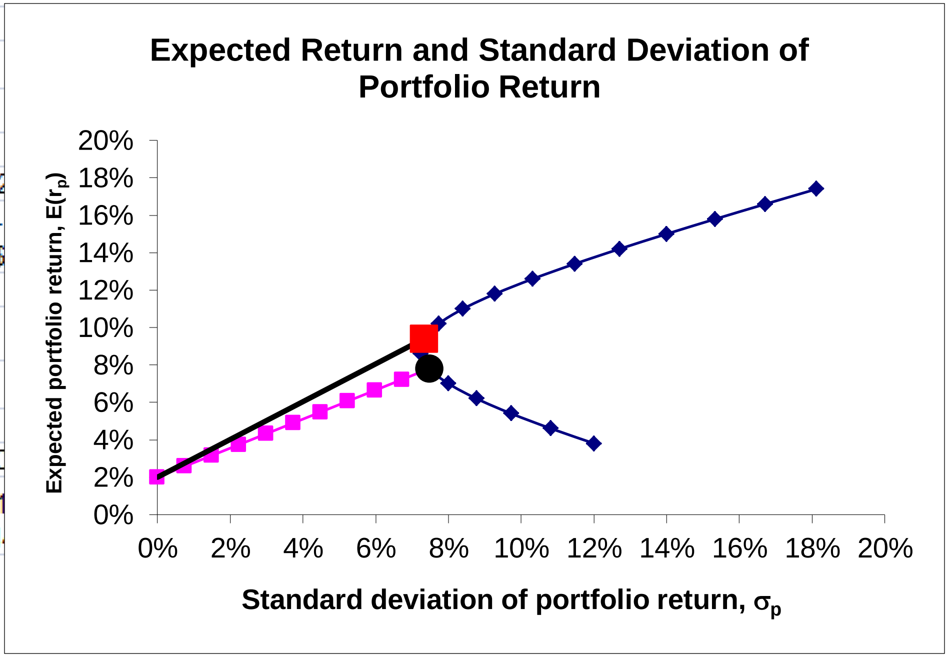
Numerical Example

	A	B	C	D	E
1		**Stock A**	**Stock B**	**Risk-free r_f**	
2	Average return	7.00%	15.00%	2%	
3	Variance	0.0064	0.0196		
4	Sigma	8.00%	14.00%		
5	Covariance of returns	0.0011			
6	Correlation	0.1000			
7	**Round dot portfolio** ●				
8	A	0.9			
9	B	0.1			
10	Portfolio mean	7.80%	<-- =B8*B2+(1-B8)*C2		
11	Portfolio standard deviation, σ_p	7.47%	<-- =SQRT(B8^2*B3+(1-B8)^2*C3+2*B8*(1-B8)*B5)		
12					



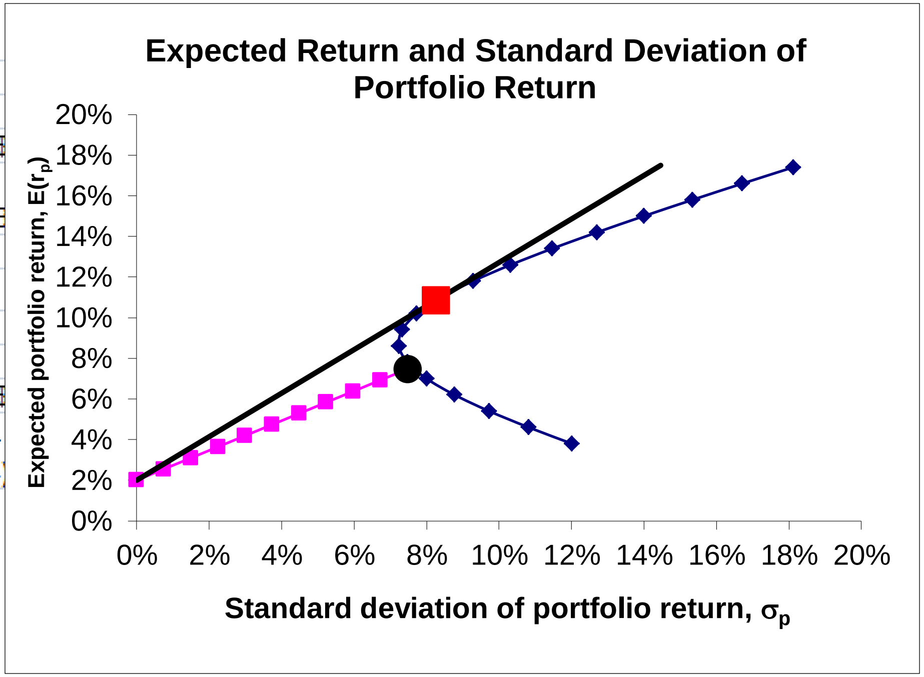
Numerical Example

	A	B	C	D	E
1		**Stock A**	**Stock B**	**Risk-free r_f**	
2	Average return	7.00%	15.00%	2%	
3	Variance	0.0064	0.0196		
4	Sigma	8.00%	14.00%		
5	Covariance of returns	0.00			
6	Correlation	0.10			
7	**Round dot portfolio** ●				
8	A	0.9			
9	B	0.1			
10	Mean	7.80%	<-- =B8*B2+(1-B8)*C2		
11	Sigma	7.47%	<-- =SQRT(B8^2*B3+(1-B8)^2*C3+2*B8*(1-B8)*$		
12					
13	**Big square portfolio**				
14	A	0.7			
15	B	0.3			
16	Mean	9.40%	<-- =B14*B2+(1-B14)*$C		
17	Sigma	7.33%	<-- =SQRT(B14^2*B3+(1-B14)^2*C3+2*B14*(1-B1		



Numerical Example

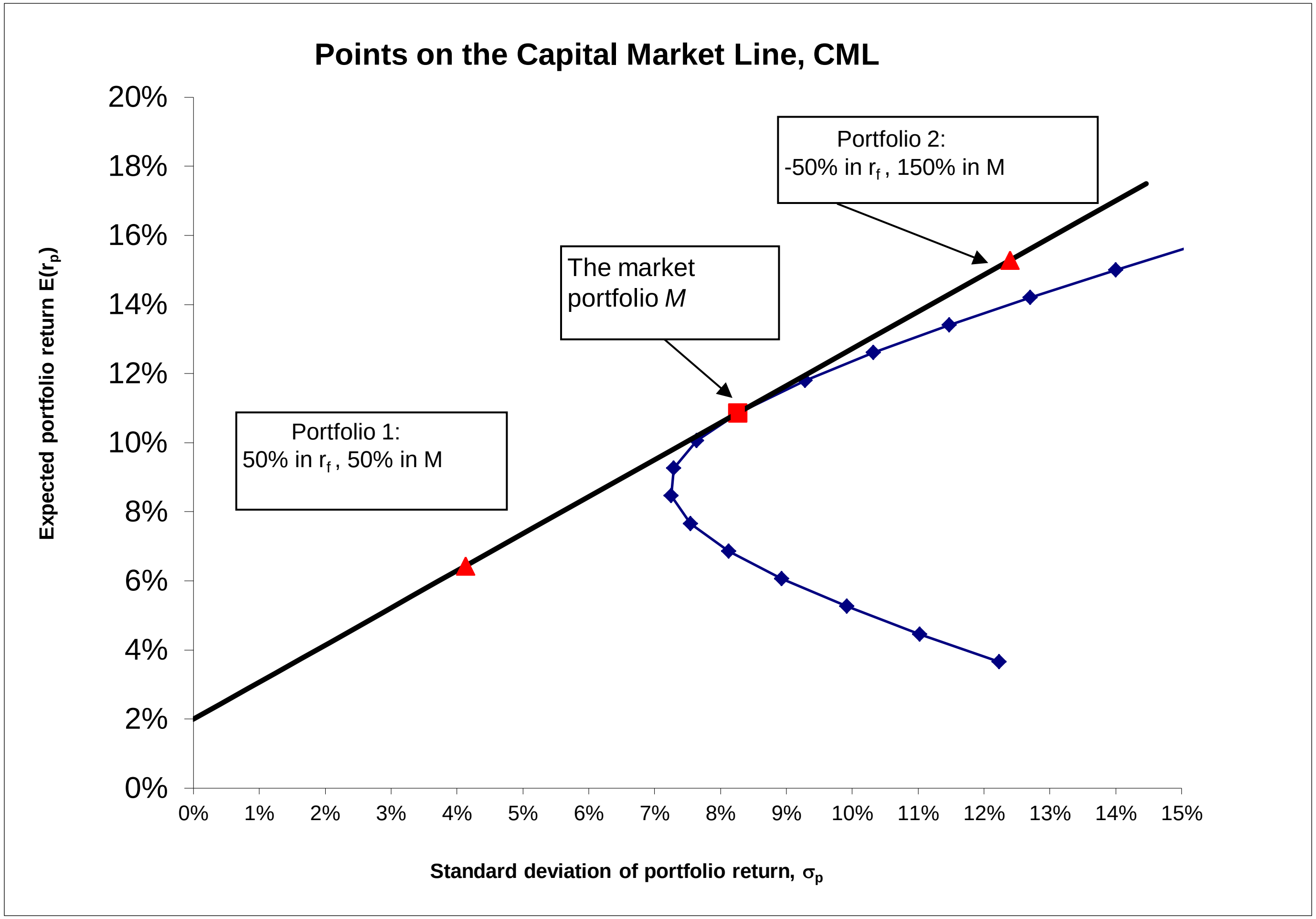
	A	B	C	D	E
1		Stock A	Stock B	Risk-free r_f	
2	Average return	7.00%	15.00%	2%	
3	Variance	0.0064	0.0196		
4	Sigma	8.00%	14.00%		
5	Covariance of returns	0.0011			
6	Correlation	0.1000			
7	Round dot portfolio ●				
8	A	0.9			
9	B	0.1			
10	Mean	7.80%	<-- =B29*B3+(1-B29)*C		
11	Sigma	7.47%	<-- =SQRT(B8^2*B3+(1-B8)^2*C3+2*B8*(1-B8)*$B		
12					
13	Best big square portfolio ■				
14	A	51.81%			
15	B	48.19%			
16	Mean	10.85%	<-- =B14*B2+(1-B14)*C		
17	Sigma	8.26%	<-- =SQRT(B14^2*B3+(1-B14)^2*C3+2*B14*(1-B14)		



About the CML

- All the portfolios on the CML incorporate this choice: Each CML portfolio is a combination of an investment in the risk-free asset Rf and the market portfolio M.
- Any portfolio on the CML is optimal in the sense that it could possibly be a rational investor's choice of his best investment portfolio.

Numerical Example



Examples of CML Portfolios

$E(r_M) = 10.85\%$, $r_f = 2\%$, and $\sigma_M = 8.26\%$.

Portfolio Proportions and Investment Returns on the Capital Market Line (CML)

Percentage invested in market portfolio *M*	$E(r_p) = \%\ in\ risk\text{-}free * r_f + \%\ in\ market * E(r_M)$	$\sigma_p = \%\ in\ market * \sigma_M$
0% (invest all your wealth in risk-free asset r_f)	$E(r_p) = 100\% * r_f = 2\%$	$\sigma_p = 0\% * \sigma_M = 0$
50% (invest 50% of your wealth in market portfolio *M* and 50% in risk-free asset)	$E(r_p) = 50\% * r_f + 50\% * E(r_M) = 50\% * 2\% + 50\% * 10.85\% = 6.43\%$	$\sigma_p = 50\% * \sigma_M = 50\% * 8.26\% = 4.13\%$
100% (invest all your wealth in market portfolio *M*)	$E(r_p) = 0\% * r_f + 100\% * E(r_M) = 100\% * 10.85\% = 10.85\%$	$\sigma_p = 100\% * \sigma_M = 100\% * 8.26\% = 8.26\%$
125% (borrow 25% of your wealth to increase investment in risky assets *M*)	$E(r_p) = -25\% * r_f + 125\% * E(r_M) = -25\% * 2\% + 125\% * 10.85\% = -0.5\% + 13.57\% = 13.06\%$	$\sigma_p = 125\% * \sigma_M = 125\% * 8.26\% = 10.33\%$
150% (borrow 50% of your wealth to increase investment in risky assets *M*)	$E(r_p) = -50\% * r_f + 150\% * E(r_M) = -50\% * 1\% + 150\% * 10.85\% = -1\% + 16.28\% = 15.28\%$	$\sigma_p = 150\% * \sigma_M = 150\% * 8.26\% = 12.39\%$
200% (borrow 100% of your wealth to increase investment in risky assets *M*)	$E(r_p) = -100\% * r_f + 200\% * E(r_M) = -100\% * 2\% + 200\% * 10.85\% = -2\% + 21.70\% = 19.70\%$	$\sigma_p = 200\% * \sigma_M = 200\% * 8.26\% = 16.52\%$

The Separation Property: task 1

- Introduced by James Tobin, the 1983 Nobel Laureate for Economics.
- It implies that portfolio choice can be separated into two independent tasks.
- **Task 1**: determining the optimal risky portfolio M (the tangent portfolio)
- Given the particular input data, <u>the best risky portfolio is the same for all clients regardless of risk aversion.</u>

The Separation Property: task 2

- The second task, construction of the complete portfolio from a risk free asset (tbills, say) and portfolio M, however, depends on personal preferences.
- Here the client is the decision maker.
- If the optimal portfolio is the same for all clients, management is more efficient and less costly-the real competition among money managers is their choice of securities.

Practical Implications (summary)

- The analyst or planner should identify what they believe will be the best performing well diversified portfolio, call it P.
- P may include funds, stocks, bonds, international and other alternative investments.
- This portfolio will serve as the starting point for all their clients.
- The planner will then change the asset allocation between the risky portfolio and “near cash” investments according to risk tolerance of client.
- The risky portfolio P may have to be adjusted for individual clients for tax and liquidity concerns if relevant and for the client’s opinions.

Computing the tangency portfolio

- We have seen the importance of the tangency portfolio.
- Now we need to show how to calculate it.
- Recall:

$$E[R_p] = r_f + w(E[R_A]\text{-}r_f)$$
$$= r_f + \sigma_p \left(\frac{E[R_A]\text{-}r_f}{\sigma_A} \right)$$

The Sharpe ratio

- The slope of the capital market line is the Sharpe ratio of the risky portfolio, that is the ratio of excess return to standard deviation:

$$S_A = \frac{E[\mathrm{R}_{\mathrm{A}}]\text{-}\mathrm{r}_f}{\sigma_A}$$

- This is the rate at which the investor can increase expected return by accepting higher portfolio volatility.

The Sharpe Ratio

- The Sharpe ratio is commonly used to rank portfolios in terms of their risk-return trade-off:

$$S_A = \frac{E[\mathrm{R}_{\mathrm{A}}]\text{-r}_f}{\sigma_A}$$

- A higher Sharpe measure indicates a better reward per unit of volatility, in other words, a more efficient portfolio.

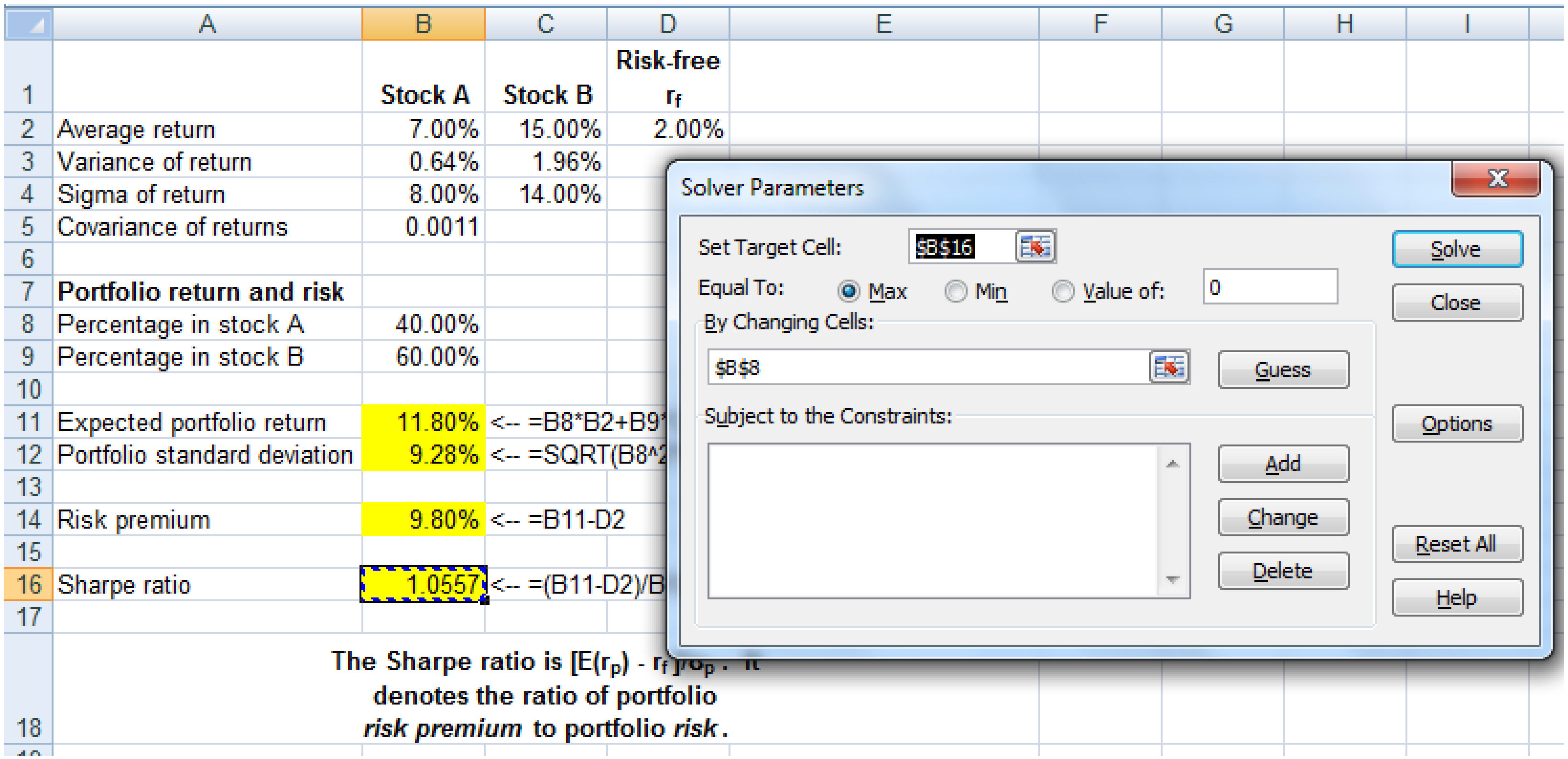
The Tangent Point

- The portfolio at the tangent point must have the highest feasible reward-to-risk ratio.
- So we find the tangent portfolio by simply finding the portfolio of assets that maximizes the Sharpe ratio.
- This is easily done using Solver in Excel, or in R.

Finding the Tangent Portfolio

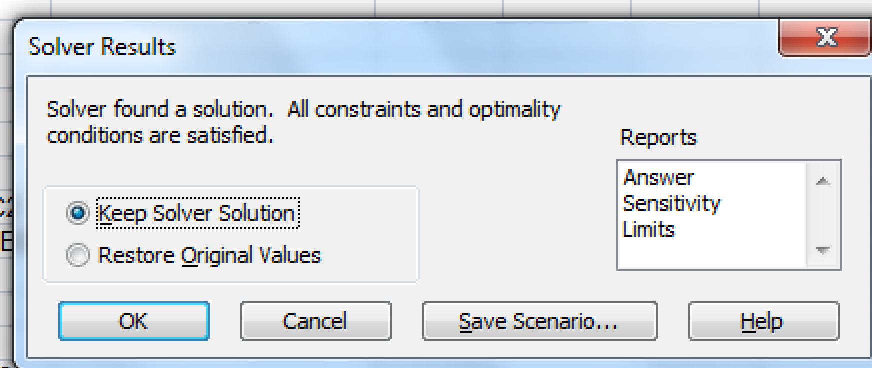
	A	B	C	D	E
1		**Stock A**	**Stock B**	**Risk-free r_f**	
2	Average return	7.00%	15.00%	2.00%	
3	Variance of return	0.64%	1.96%		
4	Sigma of return	8.00%	14.00%		
5	Covariance of returns	0.0011			
6					
7	**Portfolio return and risk**				
8	Percentage in stock A	40.00%			
9	Percentage in stock B	60.00%			
10					
11	Expected portfolio return	11.80%	<-- =B8*B2+B9*C2		
12	Portfolio standard deviation	9.28%	<-- =SQRT(B8^2*B3+B9^2*C3+2*B8*B9*B5)		
13					
14	Risk premium	9.80%	<-- =B11-D2		
15					
16	Sharpe ratio	1.0557	<-- =(B11-D2)/B12		
17					
18		**The Sharpe ratio is $[E(r_p) - r_f]/\sigma_p$. It denotes the ratio of portfolio *risk premium* to portfolio *risk*.**			
19					

Finding the tangency portfolio



The Output

	A	B	C	D	E
1		**Stock A**	**Stock B**	**Risk-free r_f**	
2	Average return	7.00%	15.00%	2.00%	
3	Variance of return	0.64%	1.96%		
4	Sigma of return	8.00%	14.00%		
5	Covariance of returns	0.0011			
6					
7	**Portfolio return and risk**				
8	Percentage in stock A	51.81%			
9	Percentage in stock B	48.19%			
10					
11	Expected portfolio return	10.85%	<-- =B8*B2+B9*C		
12	Portfolio standard deviation	8.26%	<-- =SQRT(B8^2*E		
13					
14	Risk premium	8.85%	<-- =B11-D2		
15					
16	Sharpe ratio	1.0716	<-- =(B11-D2)/B12		
17					
18		**The Sharpe ratio is $[E(r_p) - r_f]/\sigma_p$. It denotes the ratio of portfolio *risk premium* to portfolio *risk*.**			



Using R

- Finding the tangency portfolio in R
- The Setup

```
er=c(.07,.15)
asset.names=c("A","B")
names(er)=asset.names
covmat=matrix(nrow=2,ncol=2,c(.0064,.00112,.00
112,.0196))
rk.free=.02
```

Using R

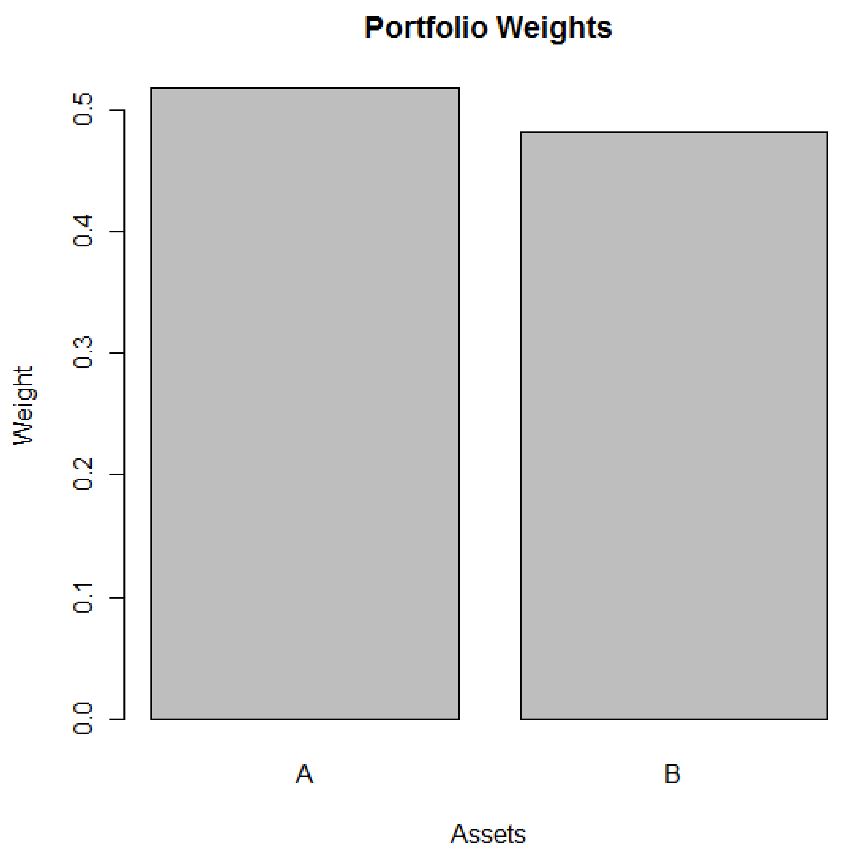
- The R code

```
tan.port <- tangency.portfolio(er, covmat, rk.free)
print(tan.port)
summary(tan.port, risk.free=rk.free)
plot(tan.port)
```

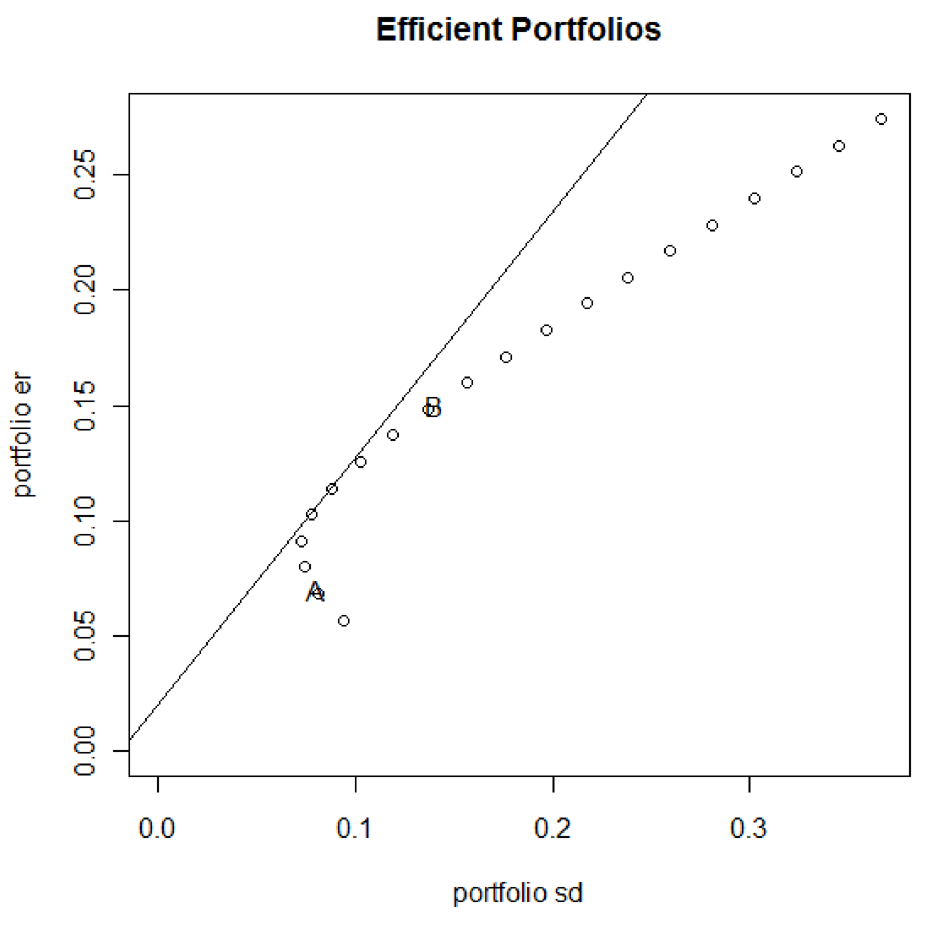
- The Output

```
Portfolio expected return:     0.1085494
Portfolio standard deviation:  0.08263454
Portfolio weights:
     A      B
0.5181 0.4819
```

The R plot



The R plot



Efficient Portfolios
portfolio er
0.00
0.05
0.10
0.15
0.20
0.25
0.0
0.1
0.2
0.3
portfolio sd
A
B

We're done with portfolios

- Or are we?
- It turns out for our three stock example, the tangency weights are as follows-is that a portfolio we want?

```
Portfolio weights:
   AAPL    RIMM     JNJ
 5.9013 -1.9240 -2.9773
```

- A lot of issues with constructing portfolios has to do with estimating the variance-covariance matrix.

R Code for 3 Stock Tangency

- This is how the weights were obtained:

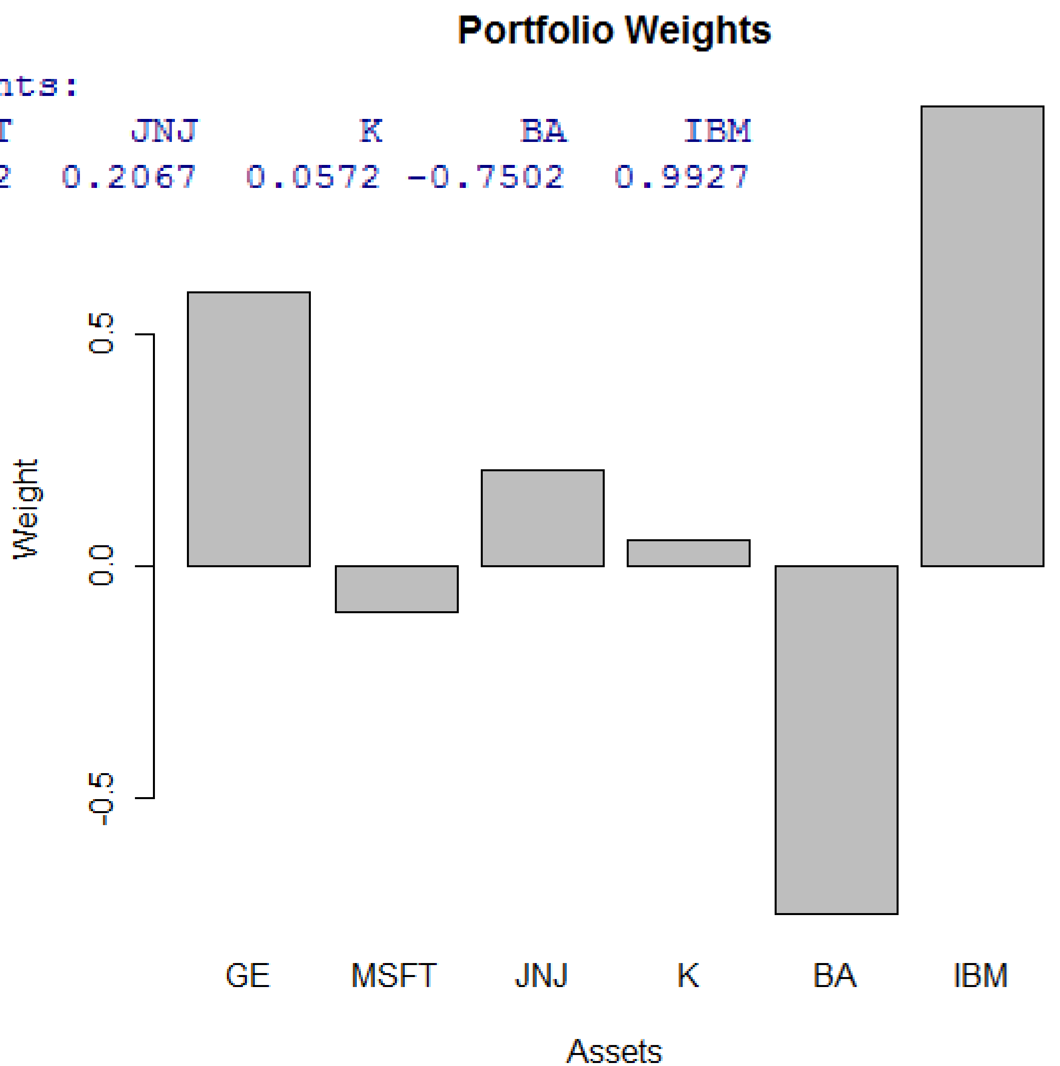
```
> tangency.portfolio(er,covmat,.001)
Call:
tangency.portfolio(er = er, cov.mat = covmat,
risk.free = 0.001)

Portfolio expected return:     0.2270952
Portfolio standard deviation:  0.5214001
Portfolio weights:
   AAPL    RIMM     JNJ
 5.9013 -1.9240 -2.9773
```

Strange Portfolio Weights (redux)

- Consider a portfolio of some big-cap stocks:
  - IBM, JNJ, GE, MSFT, K, BA and IBM
- We will construct the minimum variance portfolio of these stocks in R.
- Our results are based on yearly returns 1994-2004.

The Minimum Variance Portfolio

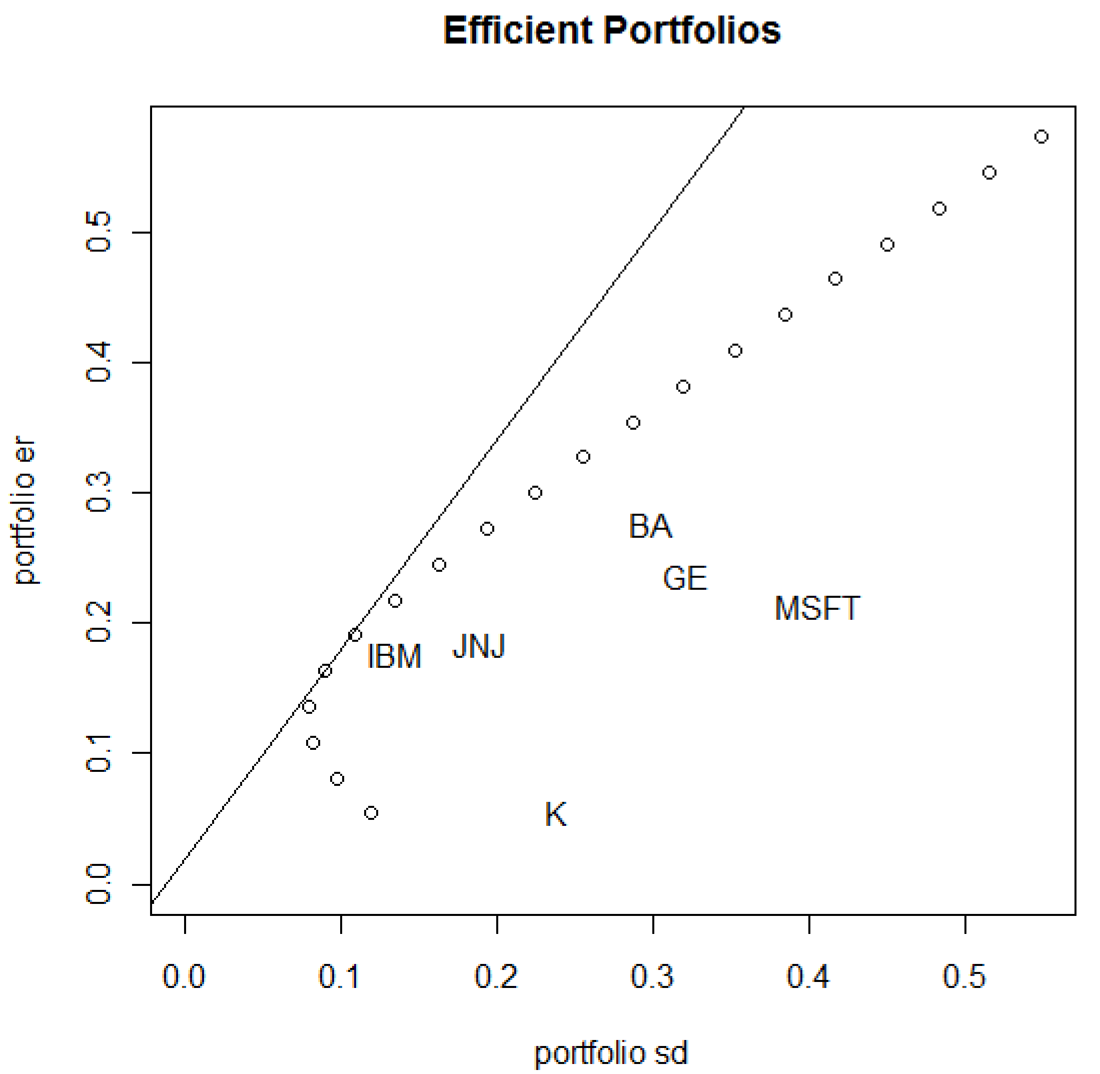
```
Portfolio weights:
     GE     MSFT      JNJ        K       BA      IBM
 0.5929 -0.0992   0.2067  0.0572 -0.7502   0.9927
```



What? The weights are a bit strange

- Note that the GMVP for our six stocks has two short positions (BA and MSFT) and that it has a very large positive position in GE and IBM.
- This is a potentially objectionable feature of computing the GMVP with the sample variance-covariance matrix: It is not credible that an investor seeking minimum variance would put 60percent of his portfolio in GE and 100 percent of his portfolio in IBM, financing these positions with a short of 75 percent in BA and 10 percent in MSFT.
- The noncredible portfolios produced by the sample variance-covariance matrix have led to a variety of techniques for alternative methods of computing this matrix, which we will discuss next time.

The Efficient Frontier



Check out the bizarre weights

```
> round(ef$weights,4)
             GE    MSFT     JNJ       K      BA     IBM
port 1  -3.4731  0.3962  2.1448 -1.1733  4.3175 -1.2122
port 2  -3.2234  0.3658  2.0258 -1.0977  4.0063 -1.0768
port 3  -2.9738  0.3354  1.9068 -1.0221  3.6951 -0.9414
port 4  -2.7241  0.3049  1.7878 -0.9466  3.3839 -0.8060
port 5  -2.4744  0.2745  1.6688 -0.8710  3.0728 -0.6706
port 6  -2.2248  0.2441  1.5498 -0.7955  2.7616 -0.5352
port 7  -1.9751  0.2137  1.4308 -0.7199  2.4504 -0.3998
port 8  -1.7254  0.1833  1.3118 -0.6444  2.1392 -0.2645
port 9  -1.4758  0.1528  1.1928 -0.5688  1.8281 -0.1291
port 10 -1.2261  0.1224  1.0738 -0.4933  1.5169  0.0063
port 11 -0.9764  0.0920  0.9548 -0.4177  1.2057  0.1417
port 12 -0.7268  0.0616  0.8357 -0.3422  0.8945  0.2771
port 13 -0.4771  0.0312  0.7167 -0.2666  0.5834  0.4125
port 14 -0.2274  0.0007  0.5977 -0.1911  0.2722  0.5479
port 15  0.0222 -0.0297  0.4787 -0.1155 -0.0390  0.6832
port 16  0.2719 -0.0601  0.3597 -0.0400 -0.3502  0.8186
port 17  0.5216 -0.0905  0.2407  0.0356 -0.6613  0.9540
port 18  0.7712 -0.1210  0.1217  0.1111 -0.9725  1.0894
port 19  1.0209 -0.1514  0.0027  0.1867 -1.2837  1.2248
port 20  1.2706 -0.1818 -0.1163  0.2622 -1.5948  1.3602
```

Discussion

- If we think that the portfolio proportions derived in this way are unrealistic, then in terms of the optimization process, there must be something wrong with either the variance-covariance matrix or the vector of means (perhaps both).
- The sample variance-covariance matrix is easily computable from historical data, but—as just shown in the previous example—it has its problems.

The Inputs to Portfolio Analysis

- As we are now probably tired of seeing, to perform a portfolio analysis one needs the mean returns of each security, the variance of return of each security, as well as knowledge of all the covariances.
- The most difficult aspect of the portfolio analysis problem is obtaining these covariances.

The traditional analysts job

- Traditionally, the job of an analyst has been to estimate the future returns of a security.
- With increasing awareness of risk, more and more analysts also produce some measure of risk as well.
- Any worthy analyst should be able to provide some uncertainty measure to their prediction of expected return.

Correlation (covariance) is different

- Correlations are another matter.
- Most firms organize their analysts along traditional industry lines.
- But for portfolio analysis one needs to know how a steel stock (for example) would relate to a drug company stock.
- Analysts aren’t really in position to be able to do this.

The computational issue

- If one is following 150-250 stocks, that is 150-250 expected mean returns to know.
- There are n(n-1)/2 correlation coefficients, so that implies 11,175 to 31,125 correlation coefficients to be estimated. It can be done but it's a bit staggering to think about.
- Recognition of this has led to the development of models to describe and predict the correlation structure-more next time.

The Matrix Approach

- We now step back to the fundamental portfolio ideas to show how to calculate our now familiar quantities for portfolio risk and return.

$$E\left[portfolio\right] = W_A * E[A] + W_B * E[B]$$

$$Var\left[portfolio\right] = W_A^2 * Var[A] + W_B^2 W[B] + 2 * W_A * W_B * Cov[A, B]$$

These equations can big quickly

- Consider four assets

$$E[portfolio] = W_A * E[A] + W_B * E[B] + W_C * E[C] + W_D * E[D]$$

$$\begin{aligned} Var[portfolio] = {} & W_A^2 * Var[A] + W_B^2 * Var[B] + W_C^2 * Var[C] \\ & + 2 * W_D^2 * Var[D] + 2 * W_A * W_B * Cov[A, B] \\ & + 2 * W_A * W_C * Cov[A, C] + 2 * W_A * W_D * Cov[A, D] \\ & + 2 * W_B * W_C * Cov[B, C] + 2 * W_B * W_D * Cov[B, D] \\ & + 2 * W_C * W_D * Cov[C, D] \end{aligned}$$

Time for matrices

- By using matrices for these calculations, a generic algorithm emerges that never changes.
- The matrix contents adjust to the number of securities under consideration.
- By producing thematr ix equivalent of these equations for the two asset case, extensions to the algorithm produce minimum variance portfolios, an efficient frontier, and a capital allocation line.

The Set-Up

- Consider the two asset case.
- We need two 2x1 vectors and a 2x2 matrix as follows:

$$M = \begin{bmatrix} E[A] \\ E[B] \end{bmatrix} \quad ; VCOV = \begin{bmatrix} Var[A] & Cov[A,B] \\ Cov[A,B] & Var[B] \end{bmatrix} \quad ; W = \begin{bmatrix} W_A \\ W_B \end{bmatrix}$$

The Matrix Math for Expected Return

- The expected return of a two asset portfolio is simply vector * vector multiplication (2x1)*(1x2) = 1x1 scalar:

$$E[portfolio] = M^T * W = \begin{bmatrix} E[A] & E[B] \end{bmatrix} * \begin{bmatrix} W_A \\ W_B \end{bmatrix}$$

$$= W_A * E[A] + W_B * E[B]$$

The Matrix Match for Portfolio Var

$$Var[portfolio] = W^T * VCOV * W$$

$$= \begin{bmatrix} W_A & W_B \end{bmatrix} * \begin{bmatrix} Var[A] & Cov[A,B] \\ Cov[A,B] & Var[B] \end{bmatrix} * \begin{bmatrix} W_A \\ W_B \end{bmatrix}$$

$$= \begin{bmatrix} W_A & W_B \end{bmatrix} * \begin{bmatrix} W_A * Var[A] + Cov[A,B] \\ W_A * Cov[A,B] + W_B * Var[B] \end{bmatrix}$$

$$= W_A^2 * Var[A] + W_A * W_B * Cov[A,B] + W_B^2 * Var[B] + W_A * W_B * Cov[A,B]$$

$$= W_A^2 * Var[A] + W_B^2 * Var[B] + 2 * W_A * W_B * Cov[A,B]$$

For those curious……

- A result about the variance-covariance matrix says that

$\mathrm{vc}(\boldsymbol{a}\boldsymbol{X}) = \boldsymbol{a}\,\mathrm{vc}(\boldsymbol{X})\boldsymbol{a}^T$ if $\boldsymbol{X}$ is a random vector in $\mathbb{R}^n$ and $\boldsymbol{a} \in \mathbb{R}^{m\times n}$.

- How does that apply here?

Finding Minimum Variance Portfolio

- A bit of an advanced idea-need to know about Lagrangian multipliers.
- Will set it up and quickly go through the math.
- Will assume a three asset portfolio

$$w = c(w_1, w_2, w_3)$$

$$\min_{w_1, w_2, w_3} \sigma_P^2 = w'\Sigma w \quad st \; w'1 = 1$$

- Analytic solution using matrix algebra
- Solver using Excel

Finding Minimum Variance Portfolios

- The Lagrangian is

$$w = c(w_1, w_2, w_3)$$
$$L(w, \lambda) = w'\Sigma w + \lambda(w'1 - 1)$$

- One that to differentiate wrt the vector of weights (there are matrix derivative results) to obtain the first order conditions.

Minimum Variance Portfolio

- One can show (see Zivot's slides on website) that the weights for the minimum variance portfolio are then given by

$$w = \frac{\Sigma^{-1}1}{1'\Sigma^{-1}1}$$

- The point being not to scare you away with the math, but to point out how matrix theory makes this easier to work with.

Example:R Code for 3 Stocks

```
er=c(0.03469, -0.01273, 0.00071)
covmat=matrix(c(0.01125, 0.01096, 0.00161,0.01096,
0.03904, 0.00204,0.00161,
0.00204, 0.00199),nrow=3,ncol=3)
names(er)=c("AAPL","RIMM","JNJ")
colnames(covmat)=c("AAPL","RIMM","JNJ")
rownames(covmat)=c("AAPL","RIMM","JNJ")

gmin.port <- globalMin.portfolio(er, covmat)
```

```
> gmin.port
Call:
globalMin.portfolio(er = er, cov.mat = covmat)

Portfolio expected return:     0.002638269
Portfolio standard deviation:  0.04438313
Portfolio weights:
   AAPL    RIMM     JNJ
 0.0511 -0.0142  0.9631
```

The Matrix Solution

```
> solve(covmat)%*%ones/as.numeric((t(ones)%*%solve(covmat)%*%ones))
            [,1]
AAPL  0.05112265
RIMM -0.01422031
JNJ   0.96309767
```

Factor Models

- In finance, people like to build regression models to mode asset returns and volatility.
- In statistics, we call the right hand side variables of a regression model "explanatory variables" or "predictor variables".
- In finance, they call the right hand side variables "**factors**".

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

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

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

$$\alpha_i, \beta_i, \sigma^2_{\varepsilon,i}, \mu_M, \sigma^2_M$$

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

- Using the formula

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

```
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