Introduction to Financial Data Analysis

Week 3-4: Portfolio Returns

Portfolio Returns and Benchmark Portfolio

- In prior sections we discussed how to calculate returns of individual securities.
- Most investors likely have more than one security in their portfolio. In those cases, we may be interested in knowing the *portfolio return*.
- **Portfolio's Return**: the weighted-average return of the securities in the portfolio with the weights given by the amount of money you invest in each security.
- Portfolio returns are used for the construction of benchmark portfolio returns, which are returns of a hypothetical portfolio that we can use to compare the performance of our portfolio.
- Example: SP500 is a benchmark portfolio for investing in US, A50 is a benchmark portfolio for investing in China A Shares.

Constructing Portfolio Returns

- The return on a portfolio of securities is equal to the weighted-average of the returns of the individual securities in the portfolio with the weight calculated as the percentage invested in that security relative to the total amount invested in the portfolio.
- Two assets example:
- $$r_p = \frac{I_1}{I_1 + I_2} * r_1 + \frac{I_2}{I_1 + I_2} * r_2 = w_1 * r_1 + w_2 * r_2,$$
- where *rp* is the portfolio return, *I*1 is the dollar amount invested in security 1, *I*2 is the dollar amount invested in security 2, *r*1 is the return of security 1, *r*2 is the return of security 2.

Constructing Portfolio Returns

- For multiple assets:
- $$r_p = w_1 * r_1 + w_2 * r_2 + w_3 * r_3 + w_4 * r_4 + \cdots = \sum_{i=1}^{N} w_i * r_i,$$
- Note that if we have cash in the portfolio, we can still assign a weight as we as a return to the cash (i.e., some short-term interest rate, such as a money market fund rate).

Benchmark Portfolio Construction

- Benchmark portfolio construction: **equal-weighted portfolio** and a **value-weighted portfolio** with monthly rebalancing.
- Equal Weighted: At the beginning of each month, invest equal amount of money in each security.
- Value Weighted: Also, Capitalization-weighted, at the beginning of each month, the amount of money invested in each security is in proportion to their market cap.
  - **Market capitalization:** refers to the total dollar **market** value of a company's outstanding shares. Commonly referred to as "**market cap**“, it is calculated by multiplying a company's shares outstanding by the current market price of one share.
  - The weight of security i is equal to the market capitalization of security i divided by the market capitalization of all securities in the portfolio.
  - Some of the major indexes use some form of value-weighting, such as the S&P 500 Index.
  - Returns of larger firms are given more weight.

Rebalancing/Reallocation

- **Rebalancing**: the process of realigning the weightings of a portfolio of assets.
  - Rebalancing involves periodically buying or selling assets in a portfolio to maintain an original or desired level of asset allocation or risk.
  - For example, say an original target asset allocation was 50% stocks and 50% bonds. If the stocks performed well during the period, it could have increased the stock weighting of the portfolio to 70%. The investor may then decide to sell some stocks and buy bonds to get the portfolio back to the original target allocation of 50/50.

Example: Equal-weighted Benchmark Portfolio Construction

- In this example, we assume the benchmark portfolio contains 10 stocks: randomly selected from CRSP database.
- The investor invest equal amount of money in each stock on the first day of each month.
- Try to calculate the equal-weighted portfolio return over a month
- The initial wealth at the t=0 is $1.
- At t=0, invest $0.1 in each stock.

Example: Equal-weighted Benchmark Portfolio Construction

- Consider stock i with returns on each date as follows

t	return	Cumulative return	$ amount on stock i
0		1	0.1
1	0.01	1.01=1*(1+0.01)	0.1*1.01
2	0.02	1.0302=1.01*(1+0.02)	0.1*1.0302
3	-0.01	1.019898=1.0302*(1-0.99)	0.1*1.019898
4	-0.03	0.9893011=1.019898*(1-0.03)	0.1*0.9893011
5	0.02	1.0090871=0.9893011*(1+0.02)	0.1*1.0090871
6	0.03	1.0393597=1.0090871*(1+0.03)	**0.1*1.0393597**

- Portfolio return: Your investment of 1 dollar gives you the summation of the highlighted number over all stocks at t=6.

Example: Equal-weighted Benchmark Portfolio Construction

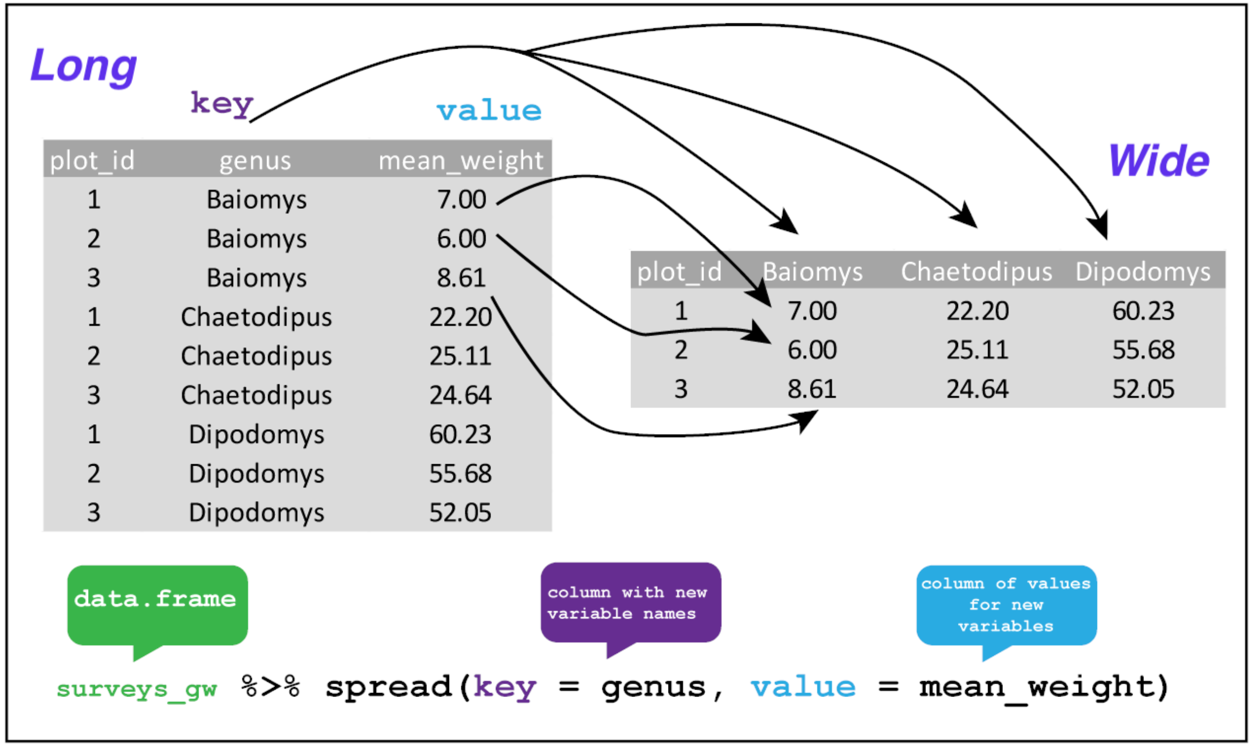
- Step 1: Grab all 10 stocks' daily returns
  - Connect to WRDS
  - Use the same random tickers selected
  - Use paste0 function to generate quote sent to the CRSP database
  - Get daily returns, and close price, shares outstanding for selected stocks
  - Save the sample data for future use.
  - Only keep permno, date, ret for this project

Example: Equal-weighted Benchmark Portfolio Construction

- Step 2: Format returns so that on each date, there are equal number of stocks in the portfolio
  - Missing stock returns should be replaced with value 0
  - The fastest way for a small amount of stocks is to use spread() and gather() function from package tidyr
  - We use the spread() function to turn the long data into a wide one
  - We use gather() function to turn the wide data back to the long data so that it can be used by the dplyr package.
  - After the two steps above, stockings with missing return on a particular date will have a return NA rather disappear from data.

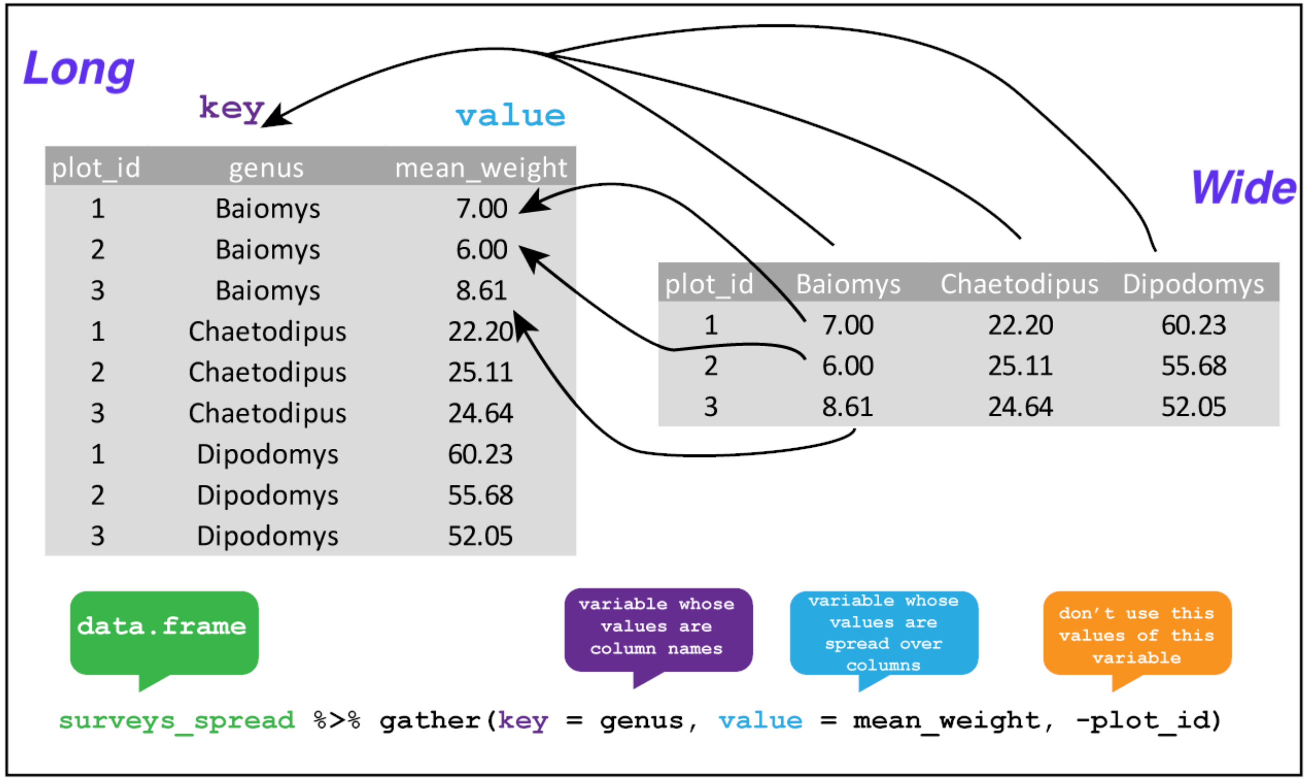
Find The Hidden Missing Returns

- **spread**() takes two columns (key & value) and spreads in to multiple columns, it makes “long” data wider.



Find The Hidden Missing Returns

- **gather**() takes two columns (key & value) and spreads in to multiple columns, it makes “long” data wider.



Example: Equal-weighted Benchmark Portfolio Construction

- Step 3. calculate the cumulative returns of each stock in each month
  - Replace returns that are NA with 0, so that within a month, each day has equal number of stocks
  - Create a label for year-month: yymm
  - Group by each stock and year-month to calculate each stock, each month's cumulative returns
- Step 4. The EW portfolio index cumulative return is the simple average of cumulative return of all stocks on each month
  - Visualize equal weight index cumulative returns for one particular month
- Step 5: Cumulative index returns over the all periods
  - Recalculate the returns of index on each date from their monthly cumulative returns
  - Recalculate the cumulative returns for the all periods.
  - Visualize equal weight index cumulative returns

Example: Value-weighted Benchmark Portfolio Construction

- We assume the benchmark portfolio contains 10 stocks: randomly selected from CRSP database.
- The amount of money an investor invest in one stock on the first day of each month is in proportion to the stock's market capital relative to the entire portfolio.
- To avoid looking forward bias, the weight we use is based on the last day from the previous month
- Try to calculate the value-weighted portfolio return over a month
- The total initial wealth at the t=0 is $1.
- At t=0, the market cap of stock i is $15 million, the entire market cap of all 10 stocks is $ 100 million, and we invest in $0.15 in stock i.

Example: Value-weighted Benchmark Portfolio Construction

- Consider stock i with returns on each date as follows

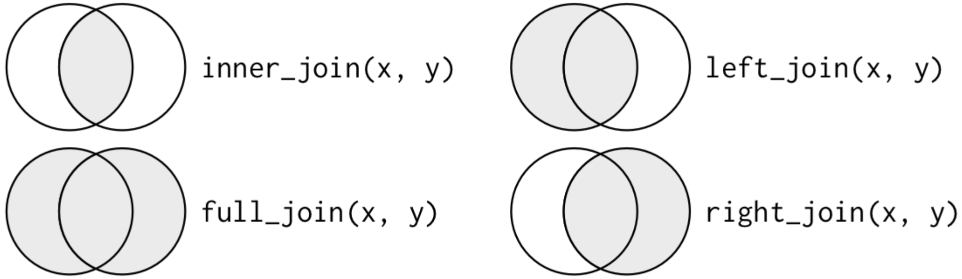
t	return	Cumulative return	$ amount on stock i
0		1	0.15
1	0.01	1.01=1*(1+0.01)	0.15*1.01
2	0.02	1.0302=1.01*(1+0.02)	0.15*1.0302
3	-0.01	1.019898=1.0302*(1-0.99)	0.15*1.019898
4	-0.03	0.9893011=1.019898*(1-0.03)	0.15*0.9893011
5	0.02	1.0090871=0.9893011*(1+0.02)	0.15*1.0090871
6	0.03	1.0393597=1.0090871*(1+0.03)	**0.15*1.0393597**

Your investment of 1 dollar gives you the summation of the highlighted number over all stocks at t=6.

Example: Value-weighted Benchmark Construction

- Step 1: grab all stocks' daily returns, close prices and shares outstanding.
- Step 2: calculate the market cap using close price*shares outstanding of each individual stock and fill returns when they are NA with 0.
  - Since the returns are market caps are processed in two separate data.frame, we rejoin them by matching their permno and date
  - left_join(), right_join(), inner_join(), full_join()
  - merge()

Join Data Using dplyr



a

x1	x2
A	1
B	2
C	3

+

b

x1	x3
A	T
B	F
D	T

=

Mutating Joins

x1	x2	x3
A	1	T
B	2	F
C	3	NA

dplyr::left_join(a, b, by = "x1")

Join matching rows from b to a.

x1	x3	x2
A	T	1
B	F	2
D	T	NA

dplyr::right_join(a, b, by = "x1")

Join matching rows from a to b.

x1	x2	x3
A	1	T
B	2	F

dplyr::inner_join(a, b, by = "x1")

Join data. Retain only rows in both sets.

x1	x2	x3
A	1	T
B	2	F
C	3	NA
D	NA	T

dplyr::full_join(a, b, by = "x1")

Join data. Retain all values, all rows.

Use Base R function merge()

- **merge**(x, y, by=c("key1","key2"), by.x=c("key1.x","key2.x"), by.y = c("key1.x","key2.y"), all.x=T/F, all.y=T/F)

dplyr	Base R
left_join(a,b,by="x1")	merge(a,b,by="x1",all.x=T,all.y=F)
right_join(a,b,by="x1")	merge(a,b,by="x1",all.x=F,all.y=T)
full_join(a,b,by="x1")	merge(a,b,by="x1",all.x=T,all.y=T)
inner_join(a,b,by="x1")	merge(a,b,by="x1",all.x=F,all.y=F)

Example: Value-weighted Benchmark Construction

- Step 3: Calculate the market cap for weighing at the beginning of each month.
  - Since the market cap used for calculating weights is the market cap of last date from last month, we first lag the market cap by 1 day, the variable is called lag.me
  - Keep the first value of lag.me for each month which is the last date me from last month. The function to get the first value in a vector is first()
  - Remove the first month of entire data as all the lag.me are NA
  - Fill the rest NA values in lag.me with 0.

Example: Value-weighted Benchmark Construction

- Step 4: calculate the weight of stock i using the market cap of stock(i)/the sum of all stocks' market cap.
- Step 5: The initial wgt of stock i is lag.me(i)/sum(lag.me(j)) .
- Step 6: Calculate the cumulative returns of each individual stock over each month
- Step 7: VW return is the sum(wgt(i)*cumulative_return(i))
- Step 8: Cumulative Returns over the all periods

Risk

- Investments inherently contain some level of risk
- What is risk? .
  - Larger swings in price are riskier than smaller swings in price.
  - More frequent price changes are riskier than less frequent price changes.
  - Government bond is less risky than corporate bonds.
- How to quantify risk?
- Markowitz gave us the most common measure of risk we use today, which is the *variance* of a security's return.
- The variance or, its positive square root, *standard deviation* is a measure of how far a security's return deviates from its average during the period.
- Shortcoming of variance as measurement of risk: it does not care whether the deviation from the average is a positive deviation or a negative deviation—both are treated as risk.

Other Measurements of Risk

- Measures on loss or downside risk: Value-at-Risk (VaR) and Expected Shortfall (also known as conditional VaR or tail VaR).
- Modifications of variance: Parkinson, Garman-Klass, Rogers-Satchell-Yoon, and Yang and Zhang. These measures have been shown to be several times more efficient than close-to-close volatility (i.e., close-to-close volatility divided by alternative risk measure).

Risk

- What is a good portfolio? Risk-return trade-off
- Individual security risk, as measured typically by the variance (or standard deviation) of returns.
- Portfolio risk: In the context of a portfolio we are less concerned with individual security variance but are more concerned with the covariance of the securities in the portfolio
- Two measurements of portfolio loss: Value-at-Risk (VaR) and Expected Shortfall (ES).

Risk-Return Trade-Off

- The main trade-off we have to consider when making investments is between risk and return.
- The only way you can get a higher expected return from a security is if you are willing to take on more risk.
- Example: Stocks and bonds' performance over the last 50 years.
- Professor Kenneth French's data, which is available from the following URL:
- http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Stock and Bond Return

- Step 1: Import Fama-French Data from Professor Kenneth French's Website. We retrieve Fama-French monthly data and save the text file (compressed initially in a ZIP file) as F-F_Research_Data_Factors.txt
  - The data starts on the fourth line.
  - The data is of fixed-width.
  - The first column of dates has six characters to the last character, the second column labeled "Mkt-RF" is eight characters after that, the third column "SMB" is eight characters after, the fourth column "HML" is eight characters after, and the last column "RF" is eight characters after. To import this fixed-width file, we use the read.fwf command with widths of 6, 8, 8, 8, and 8 characters while skipping the first four lines of the data.

Stock and Bond Return

- Step 2: Clean up Data
  - The bottom part of the file contains summary annual data, which we do not need. As such, we delete the data from Row 1117 to Row 1213. Note that since the data on Professor French's website gets updated periodically, a later version of the file will likely end up with different row numbers.
  - We delete the corresponding rows
  - Rename V1 to V5 to something more meaningful.
  - The variables are all Factor variables and we need to convert the data to numeric variables. For the date variable,
  - Change the text.date to date variable

Stock and Bond Return

- Step 3: Calculate Raw Market Return
  - The Fama-French data gives us the excess market return, which is the return on the market less the risk-free rate.
  - Add the risk-free rate back to get the raw market return
- Step 4: Subset Data from December 1968 to December 2018
- Note that the returns are at the end of each month, so to get 50 years of returns we would assume an investment made at the end of December 1968, which is effectively the start of January 1969.

Stock and Bond Return

- Step 5: Calculate Gross Returns for the Market and Risk-free Rate. We also set the return for December 1963 to 0.
- Step 6: Calculate Cumulative Returns for the Market and Risk-free Rate Using the cumprod command, we can cumulate the market return and risk-free rate through the end of 2018.
- Step 7: Plot Stock and Bond Cumulative Performance
- Step 8: Plot Stock and Bond Returns as demonstration of their volatilities

Stock and Bond Return

- Figure clearly shows the larger returns of investing in stocks compared to investing in bonds (technically, the data shows T-bills but we will be a little loose with the terminology at this point).
- This 10× difference in value at the end of 50 years is even more impressive when we consider this difference includes the effects of the 1987 crash, the recession in the early 2000s, and the 2008/2009 crisis.
- Given how much stocks outperformed bonds over the last 50 years, why then would investors bother putting money in bonds?
- The reason for this is that an investment in stocks is riskier than an investment in bonds.

Individual Security Risk

- It is common for investors to use variance or, its positive square root, standard deviation as the measure of risk.
- The variance of an asset's return is

$$\sigma^2 = \frac{1}{T-1}\sum_{t=1}^{T}(R_t - \bar{R})^2,$$

- where *Rt* is the return of the asset on day *t* for *t* = 1, . . . , *T* and Rbar is the average return of the asset over days 1 to *T* .
- The standard deviation is equal to the square root of the variance and is denoted by *σ.*

Example: Stocks Risk and Returns over Years

- Summary of this chapter
- Calculate the multiple stocks' volatilities in each year and over the entire periods in one go.
- Tickers of selected stocks: "AAPL","MSFT","IBM","BABA","GE","KO"
- Step 1: connect to WRDS
- Step 2: get permno from DSENAMES
- Step 3: get price from DSF
- Step 4: 1. group stock by permno,
- 2. calc daily returns
- 3. calc overall periods' standard deviation
- 4. create year label
- 5. group by permno and year
- 6. calc each stock, each year's volatility and average returns

Portfolio Risk

- Markowitz (1952) shows that the covariance of the assets in the portfolio is important when assessing risk in the context of a portfolio.
- Diversification: When we add assets that are not perfectly correlated with the securities in our portfolio, the overall risk of the portfolio decreases.
- A portfolio’s risk is not simply the weighted average of the standard deviations of each of the securities in the portfolio but likely something lower.

Two Assets (Manual Approach)

- A two-asset portfolio, the portfolio risk is calculated as

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_1 w_2 \sigma_{12}$$

where w_i is the weight of security i, σ_i^2 is the variance of security i, σ_{12} is the covariance between security 1 and 2

Two Asset Example: AAPL and MSFT

- Let us work through an example of calculating the portfolio risk for a $ 10,000 portfolio that invested $ 2500 in AAPL and $ 7500 in MSFT.
- Calculate Weights of Securities in the Portfolio: the weights for our securities would be 25% AAPL and 75% MSFT.
- Calculate Standard Deviation and Covariance of the Securities
  - Calculate standard deviation using the sd command.
  - Calculate covariance using the cov command.
  - The Annual Sd = Std calculated using daily return * sqrt(252)
  - The Annual Cov or Variance = Cov or Varaince calculated using daily return * 252

Two Assets (Matrix Algebra)

- The portfolio risk in a matrix format is calculated as

$$\sigma_P^2 = w \Sigma w^T,$$

- where *w* is a row vector of weights and *Σ* is the Variance Covariance matrix of the securities in the portfolio.
- In the previous example, the w = [0.25,0.75]

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

Two Asset Example: AAPL and MSFT

- Step 1: Create a matrix of 1*2 dimension of weights
- Step 2: Construct Variance–Covariance Matrix
- Step 3: Calculate Portfolio Risk
  - To multiply matrices, we use %*% or the asterisk (*) symbol sandwiched in between two percent (%) symbols
  - To transpose a matrix, use t()

Portfolio Risk: Multiple Assets

- We use 6 stocks:
  - Tickers of selected stocks:"AAPL","MSFT","IBM","BABA","GE","KO”
  - Invest AAPL 15%, MSFT 20%, IBM 10%, BABA 25%, GE 10%, KO 20% correspondingly
- Step 1: get their return from WRDS
- Step 2: Modify dataframe structure from long to a wide one using spread() function from library(tidyr)
  - To keep names of stocks intuitive, add back tickers by merge the tickers with the returns data on “permno”
- Step 3: Calculate covariance matrix
  - Covariance matrix is automatically arranged alphabetically
- Step 4: Create a vector of weights
  - Input the wgt data together with ticker to avoid miss matching mistakes
- Step 5: Rearrange the weights by ticker alphabetically, as in the covariance matrix
  - Use order() function
- Step 6: Calculate the var and sd of the portfolio using the matrix expression

$$\sigma_P^2 = w \Sigma w^T,$$

Asset Allocation: Mean-Variance Efficient Portfolio

- We have shown that when two assets are not perfectly correlated, investing in both assets can reduce portfolio riskness.
- Questions: What is the optimal weights to allocate your investment so that your portfolio risk is smallest for your return target.
- Asset Allocation: Implementation of an investment strategy that attempts to balance risk versus reward by adjusting the weight of each asset in an investment portfolio according to the investor's risk tolerance, goals and investment time frame.
- A portfolio that achieves the smallest risk for a given return target is called a Mean-Variance Efficient Portfolio

Asset Allocation: Two Assets

- Consider two assets:
- The investment opportunity set is defined by:

$$r_p = w_i r_i + w_j r_j$$

$$\sigma_p = \sqrt{w_i^2 \sigma_i^2 + 2 w_i w_j \sigma_{ij} + w_j^2 \sigma_j^2}$$

$$w_i + w_j = 1$$

Asset Allocation: Two Risky Asset

Annual %	Equity	Bond
Return	6.65	3.06
Standard Dev	15.11	7.72
Correlation	16.70	
Covariance	0.016	

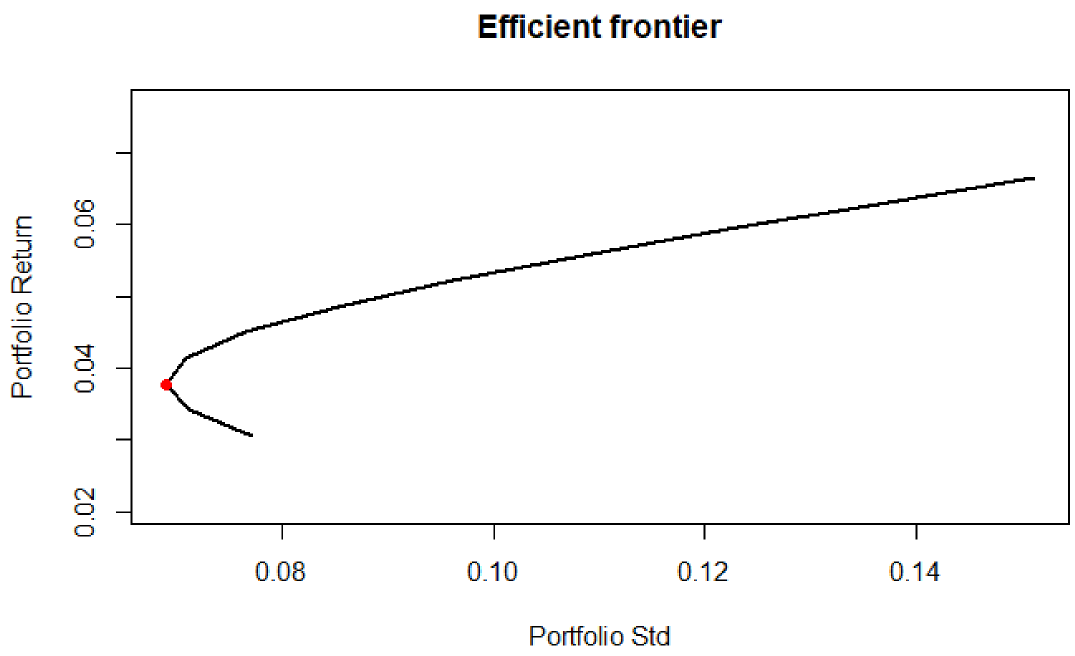
Investment opportunity set: Efficient frontier

wgt.eq	wgt.bond	sd.port	ret.port
0	1	7.72%	3.06%
0.1	0.9	7.13%	3.42%
0.2	0.8	6.91%	3.78%
0.3	0.7	7.10%	4.14%
0.4	0.6	7.67%	4.50%
0.5	0.5	8.53%	4.86%
0.6	0.4	9.62%	5.22%
0.7	0.3	10.86%	5.58%
0.8	0.2	12.21%	5.93%
0.9	0.1	13.63%	6.29%
1	0	15.11%	6.65%

Efficient Frontier

Choose the corresponding weights on the efficient frontier based on your investment target return or risk

- The Red Point is the global minimum variance portfolio (GMV).



Diversification

- Previous we have shown that when two assets are not perfectly correlated, the portfolio risk is lower than single asset risk.
- What determines the GMV point: individual asset's std and correlation
- Question: How does the correlation between two assets affect the portfolio risk exactly?
- The correlation and covariance between two assets has the relationship:

$$\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$$

- Correlation is a number between -1 and 1

Diversification

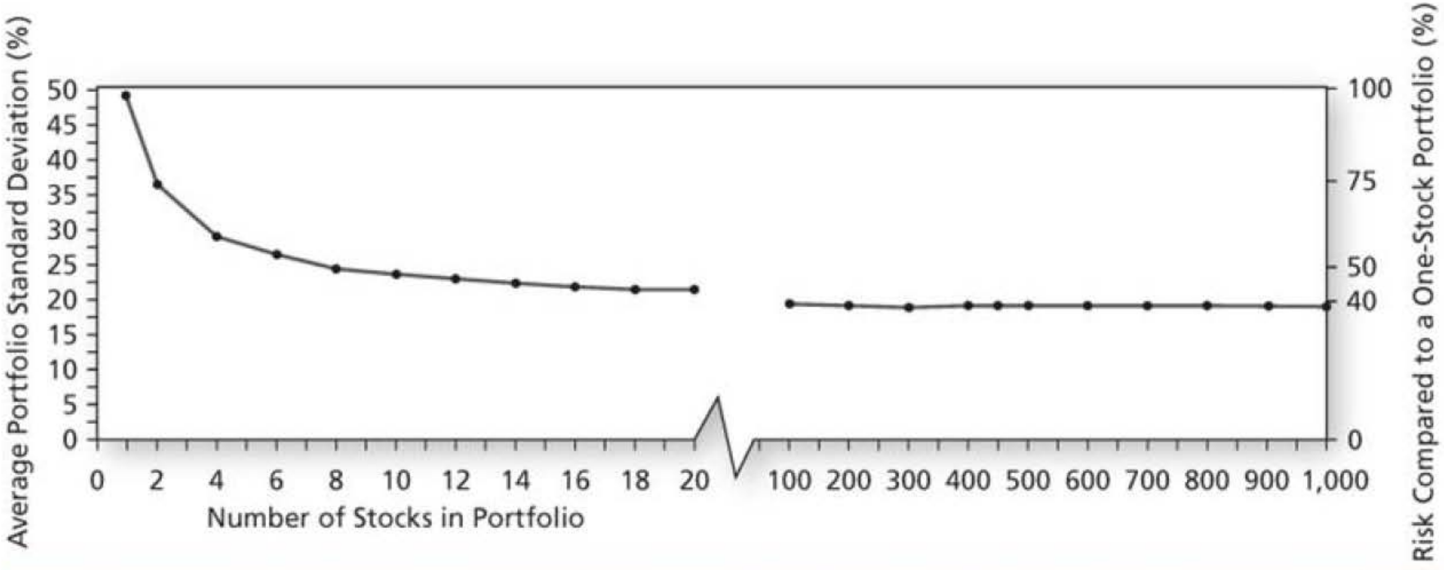
- Consider allocating your wealth between the following two assets
- Conclusion: Portfolio risk increases with correlation between the two assets

$\sigma_i = 10\%, \quad \sigma_j = 5\%, \quad w_i = w_j = 0.5$

ρ_{ij}	σ_{ij}	σ_p^2	σ_p
-1.0	-0.0050	0.0006	2.50%
-0.05	-0.0025	0.0019	4.33%
0.0	0.0000	0.0031	5.59%
0.5	0.0025	0.0044	6.61%
1.0	0.0050	0.0056	7.50%

Diversification

- In the market, the correlation can hardly be 1.
- The average standard deviation of returns of portfolios composed of only one stock is about 50%.
- The average portfolio risk falls rapidly as the number of stocks included in the portfolio increases.
- In the limit, the portfolio risk can be reduced to only 20%.



Source: From Meir Statman, "How Many Stocks Make a Diversified Portfolio?" *Journal of Financial and Quantitative Analysis* 22 (September 1987). Reprinted by permission.

Implication of Diversification

- Implications of diversification for asset pricing:
- Idiosyncratic risk is firm specific (CEO passing, new product) and can be eliminated through diversification.
- Systematic risk is common to the entire market and cannot be diversified away.

Complete Portfolio with Risk-free Asset

- Complete portfolio composed of the risk-free asset and your risky portfolio (e.g., equity, bond, or a portfolio consisting of equity and bond)
- Investment opportunity sets with risk-free asset is

$$r_c = w_{rf} r_f + w_p r_p$$

$$\sigma_c = \sqrt{w_{rf}^2 \sigma_f^2 + 2 w_p w_p \sigma_{rf,p} + w_p^2 \sigma_p^2} \Rightarrow \sigma_c = w_p \sigma_p$$

$$w_{rf} + w_p = 1 \Rightarrow w_{rf} = 1 - w_p$$

Complete Portfolio with Risk-free Asset

Substitute w_{rf} by $1 - w_p$:

$$r_c = (1 - w_p) r_f + w_p r_p$$

$$\Rightarrow w_p = \frac{r_c - r_f}{r_p - r_f}$$

In the equation of portfolio standard deviation, substitute w_p by $\dfrac{r_c - r_f}{r_p - r_f}$:

$$\sigma_c = \frac{r_c - r_f}{r_p - r_f}\sigma_p, \Rightarrow r_c = r_f + \frac{r_p - r_f}{\sigma_p}\sigma_c$$

Complete Portfolio with Risk-free Asset

$r_c = r_f + \frac{r_p - r_f}{\sigma_p}\sigma_c$ shows that return on a complete portfolio is linearly related to its standard deviation.

- Sharpe Ratio: $\frac{r_p - r_f}{\sigma_p}$
- Sharpe Ratio measures the return rewards for per unit of standard deviation.

Risk-free asset and Risky Portfolio

Asset Allocation: Two Risky Asset

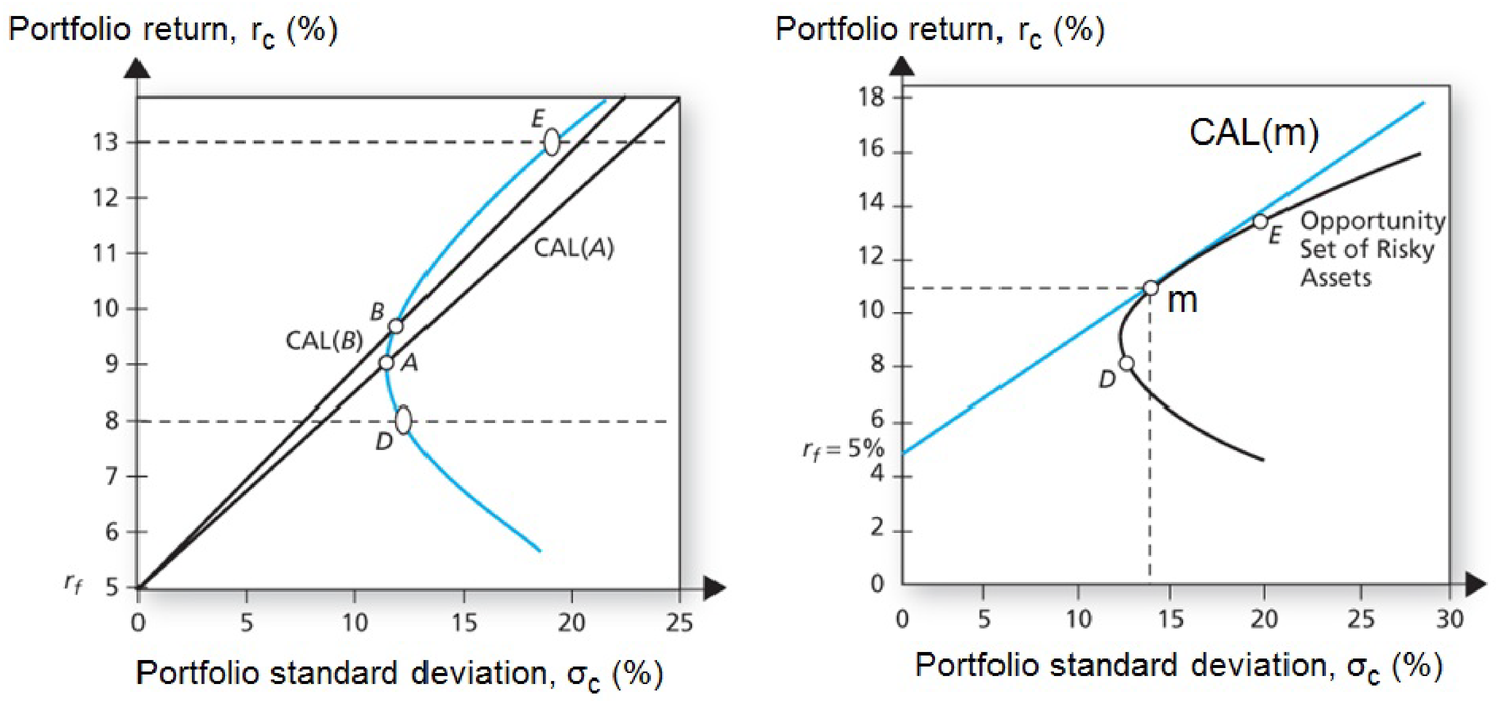
Annual %	Risk-free	Risky Portfolio (50% Equity, 50% Bond)
Return	2.00	4.86
Standard Dev	0	8.53
Correlation	0	
Covariance	0	

Capital Allocation Line (CAL)

- The investment opportunity set of complete portfolios is known as Capital Allocation Line (CAL)
- Step 1: Get the risky portfolio return and std
  - The risky portfolio has wgt = 0.4 invested in equity
- Step 2: Create a dataframe.
  - The weights allocated to the risk-free asset changes from -1 to 1.
  - When the weights of risky-free asset is negative, the investor is borrowing money to buy the risky asset
  - Calculate the corresponding return and std of the portfolio with the risk-free asset
- Step 3: Visualize and calculate the Sharpe ratio

Capital Market Line (CML)

- How to allocate two risky assets when there exist a risk free asset
- Choose the portfolio that gives the highest Sharpe Ratio
- The most efficient investment opportunity set (CAL on the right panel) is known as the Capital Market Line (CML)



Find the optimal allocation with a risk-free asset: CML

- The CML line has the highest sharpe ratio among all CAL lines
- Step 1: calculate the sharpe ratio for different combinations of equity and bond allocation
- Step 2: Find the allocation with the highest sharpe ratio