

# Financial Econometrics

## Lecture Notes

Professor Doron Avramov  
The Hebrew University of Jerusalem  
& Chinese University of Hong Kong

# Introduction: Why do we need a course in Financial Econometrics?

# Syllabus: Motivation

The past few decades have been characterized by an extraordinary growth in the use of quantitative methods in the analysis of various asset classes; be it equities, fixed income securities, commodities, and derivatives.

In addition, financial economists have routinely been using advanced mathematical, statistical, and econometric techniques in a host of applications including investment decisions, risk management, volatility modeling, interest rate modeling, and the list goes on.

# Syllabus: Objectives

- This course attempts to provide a fairly deep understanding of such techniques.
- The purpose is twofold, to provide research tools in financial economics and comprehend investment designs employed by practitioners.
- The course is intended for advanced master and PhD level students in finance and economics.

# Syllabus: Prerequisite

- I will assume prior exposure to matrix algebra, distribution theory, Ordinary Least Squares, Maximum Likelihood Estimation, Method of Moments, and the Delta Method.
- I will also assume you have some skills in computer programming beyond Excel.
  - MATLAB and R are the most recommended for this course. OCTAVE could be used as well, as it is a free software, and is practically identical to MATLAB when considering the scope of the course.
  - If you desire to use STATA, SAS, or other comparable tools, please consult with the TA.

# Syllabus: Grade Components

- Assignments (36%): there will be two problem sets during the term. You can form study groups to prepare the assignments - up to three students per group. The assignments aim to implement key concepts studied in class.
- Class Participation (14%) - Attending all sessions is mandatory for getting credit for this course.
- Final Exam (50%): based on class material, handouts, assignments, and readings.



# Syllabus: Topics to be Covered - #1

## Overview:

- Matrix algebra
- Regression analysis
- Law of iterated expectations
- Variance decomposition
- Taylor approximation
- Distribution theory
- Hypothesis testing
- OLS
- MLE

# Syllabus: Topics to be Covered - #2

- Time-series tests of asset pricing models
- The mathematics of the mean-variance frontier
- Estimating expected asset returns
- Estimating the covariance matrix of asset returns
- Forming mean variance efficient portfolio, the Global Minimum Volatility Portfolio, and the minimum Tracking Error Volatility Portfolio.



# Syllabus: Topics to be Covered - #3

- The Sharpe ratio: estimation and distribution
- The Delta method
- The Black-Litterman approach for estimating expected returns.
- Principal component analysis.

# Syllabus: Topics to be Covered - #4

- Risk management and downside risk measures: value at risk, shortfall probability, expected shortfall (also known as C-VaR), target semi-variance, downside beta, and drawdown.
- Option pricing: testing the validity of the B&S formula
- Model verification based on failure rates

# Syllabus: Topics to be Covered - #5

- Predicting asset returns using time series regressions
- The econometrics of back-testing
- Understanding time varying volatility models including ARCH, GARCH, EGARCH, stochastic volatility, implied volatility (VIX), and realized volatility

# Session #1 – Overview

# Let us Start

This session is mostly an overview. Major contents:

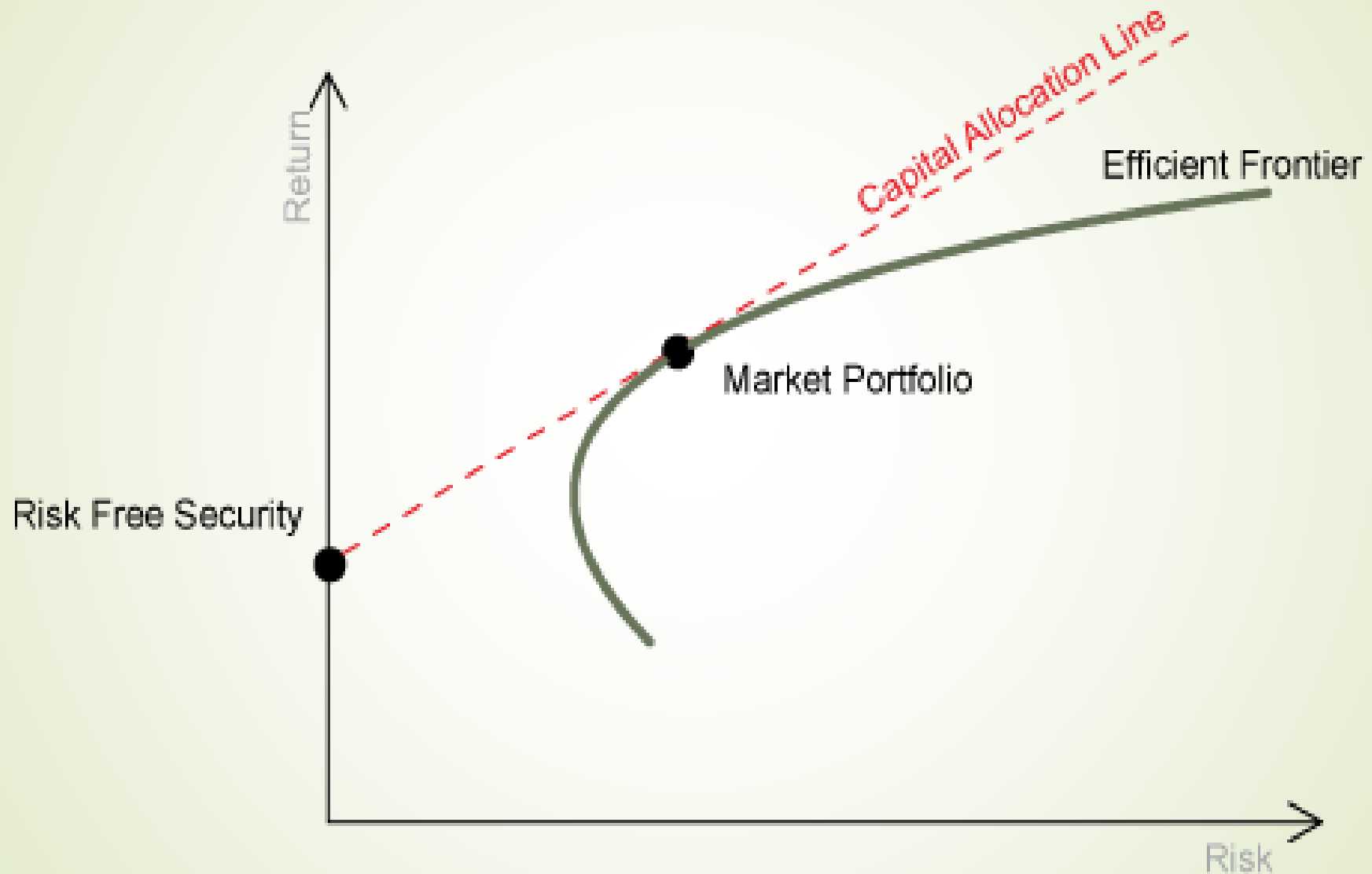
- Why do we need a course in financial econometrics?
- Normal, Bivariate normal, and multivariate normal densities
- The Chi-squared, F, and Student t distributions
- Regression analysis
- Basic rules and operations applied to matrices
- Iterated expectations and variance decomposition

# Financial Econometrics

- In previous courses in finance and economics you had mastered the concept of the efficient frontier.
- A portfolio lying on the frontier is the highest expected return portfolio for a given volatility target.
- Or it is the lowest volatility portfolio for a given expected return target.



# Plotting the Efficient Frontier



# However, how could you

## Practically Form an Efficient Portfolio?

- Problem: there are far TOO many parameters to estimate.  
For instance, investing in ten assets requires:

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \mu_{10} \end{bmatrix} \quad \begin{bmatrix} \sigma_1^2 \dots \dots \dots \sigma_{1,10} \\ \sigma_{1,2}, \sigma_2^2 \dots \dots \dots \\ \dots \dots \dots \dots \dots \dots \\ \dots \dots \dots \dots \dots \dots \\ \sigma_{1,10} \dots \dots \dots \sigma_{10}^2 \end{bmatrix}$$

which is about ten estimates for expected return, ten for volatility, and 45 for co-variances/correlations.

- Overall, 65 estimates are required.
- That is a lot given the limited amount of data.

More generally,

if there are  $N$  investable assets, you need:

- $N$  estimates for means,
- $N$  estimates for volatilities,
- $0.5N(N-1)$  estimates for correlations.

Overall:  $2N+0.5N(N-1)$  estimates are required!

Mean, volatility, and correlation estimates are noisy as reflected through their standard errors.

Beyond such parameter uncertainty there is also model uncertainty.

It is about the uncertainty about the correct model underlying the evolution of expected returns, volatilities, and correlations.

# Sample Mean and Volatility

➤ The volatility estimate is typically less noisy than the mean estimate (more later).

➤ Consider  $T$  asset return observations:

$$R_1, R_2, R_3, \dots, R_T$$

➤ When returns are IID, the mean and volatility are estimated as

$$\bar{R} = \frac{\sum_{t=1}^T R_t}{T} \quad \text{Less Noise} \rightarrow \hat{\sigma} = \sqrt{\frac{\sum_{t=1}^T (R_t - \bar{R})^2}{T - 1}}$$

# Estimation Methods

- One of the ideas here is to introduce various methods in which to estimate the comprehensive set of parameters.
- We will discuss asset pricing models and the Black-Litterman approach for estimating expected returns.
- We will further introduce several methods for estimating the large-scale covariance matrix of asset returns.

# Mean-Variance vs. Down Side Risk

- We will comprehensively cover topics in mean variance analysis.
- We will also depart from the mean variance paradigm and consider down side risk measures to form as well as evaluate investment strategies.
- Why should one resort to down side risk measures?



# Down Side Risk

- For one, investors typically assign higher weight to the downside risk of investments than to upside potential.
- The practice of risk management as well as regulations of financial institutions are typically about downside risk – such as VaR, shortfall probability, and expected shortfall.
- Moreover, there is a major weakness embedded in the mean variance paradigm.

# Drawback in the Mean-Variance Setup

- To illustrate, consider two risky assets (be it stocks) A and B.
- There are five states of nature in the economy.
- Returns in the various states are described on the next page.

# Drawback in Mean-Variance

	S1	S2	S3	S4	S5
A	5.00%	8.25%	15.00%	10.50%	20.05%
B	3.05%	2.90%	2.95%	3.10%	3.01%

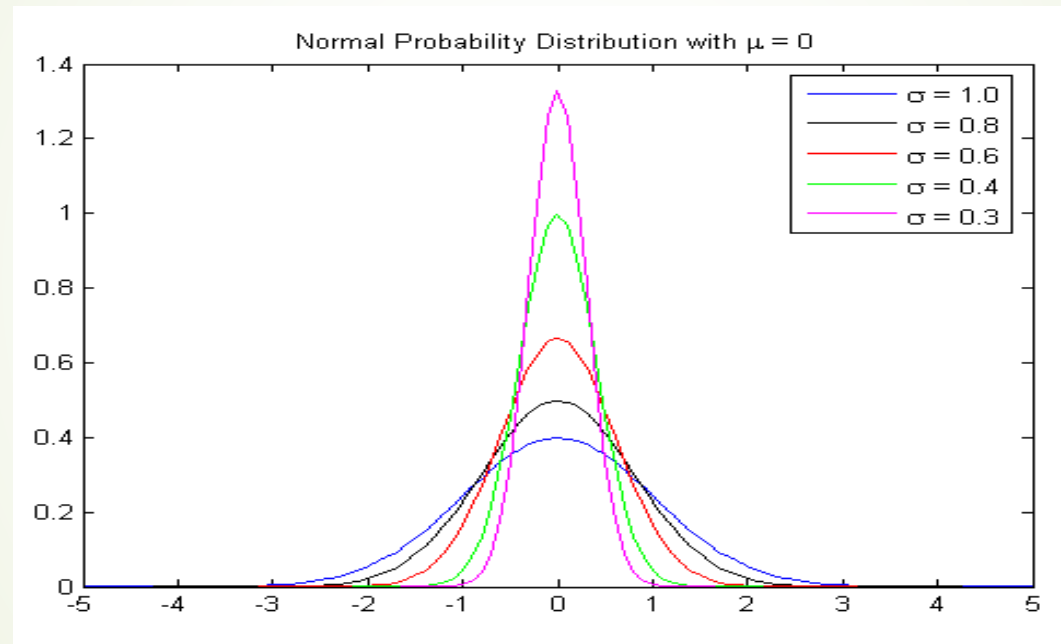
- Stock A dominates stock B in every state of nature.
- Nevertheless, a mean-variance investor may consider stock B because it may reduce the portfolio's volatility.
- Investment criteria based on down size risk measures could get around this weakness.

# The Normal Distribution

- In various applications in finance and economics, a common assumption is that quantities of interests, such as asset returns, economic growth, dividend growth, interest rates, etc., are normally (or log-normally) distributed.
- The normality assumption is primarily done for analytical tractability.
- The normal distribution is symmetric.
- It is characterized by the mean and the variance.

# Dispersion around the Mean

- Assuming that  $x$  is a zero mean random variable.
- As  $x \sim N(0, \sigma^2)$  the distribution takes the form:



- When  $\sigma$  is small (big) the distribution is concentrated (dispersed)

# Probability Distribution Function

- The Probability Density Function (pdf) of the normal distribution for a random variable  $r$  takes the form

$$pdf(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \frac{(r - \mu)^2}{\sigma^2} \right]$$

- Note that  $pdf(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}}$ , and further  
if  $\sigma = 1$ , then  $pdf(\mu) = \frac{1}{\sqrt{2\pi}}$

- The Cumulative Density Function (CDF) is the integral of the pdf, e. g.,  $cdf(\mu) = 0.5$ .



# Probability Integral Transform

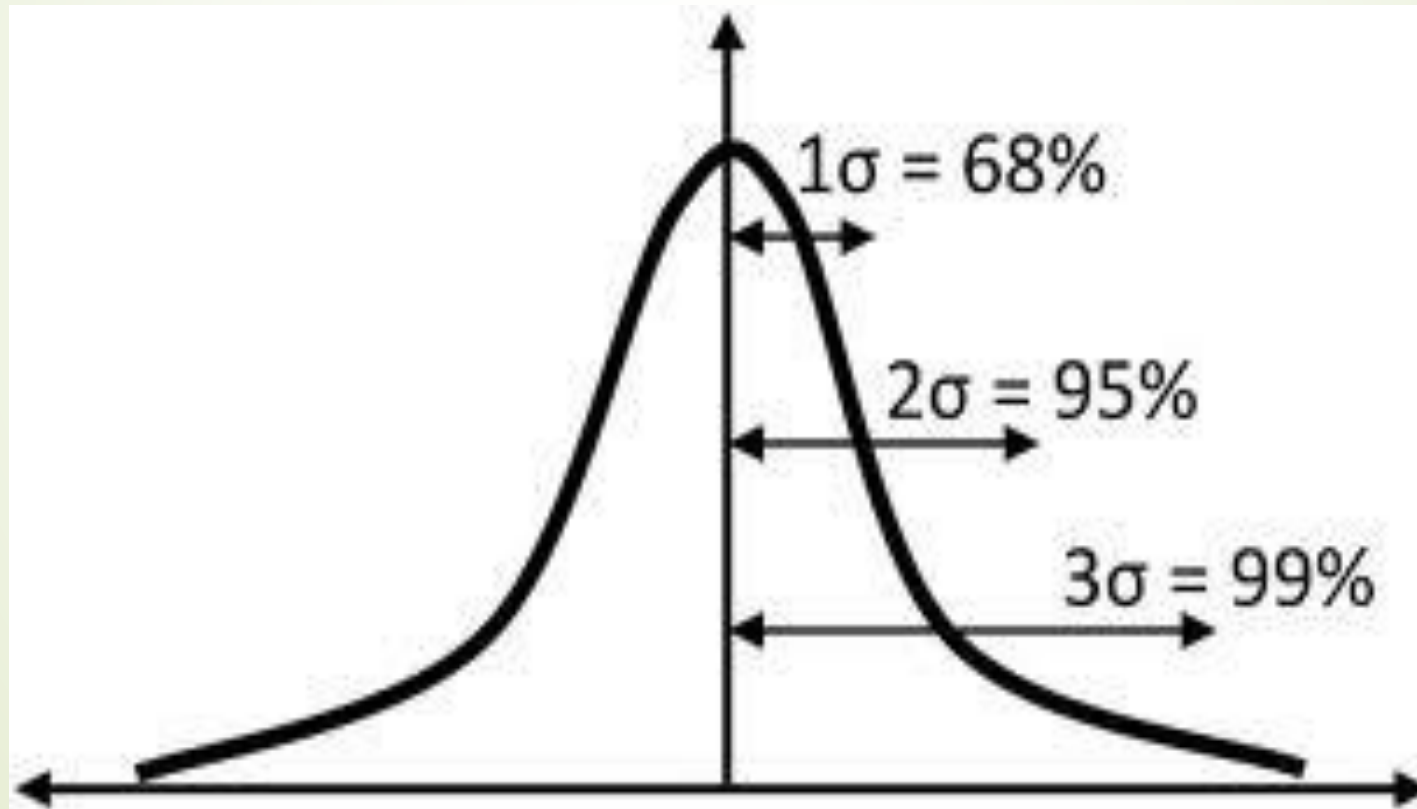
- Assume that  $x$  is normally distributed – what is the distribution of  $y=F(x)$ , where  $F$  is a cumulative density Function?
- Could one make a general statement here?

# Normally Distributed Return

- Assume that the US excess rate of return on the market portfolio is normally distributed with annual expected return (equity premium) and volatility given by 8% and 20%, respectively.
- That is to say that with a nontrivial probability the *realized return* can be negative. See figures on the next page.
- The distribution around the *sample mean return* is also normal with the same expected return but smaller volatility.

# Confidence Level

Normality suggests that deviation of 2 SD away from the mean creates an 80% range (from -32% to 48%) for the realized return with approx. 95% confidence level (assuming  $r \sim N(8\%, (20\%)^2)$ )



# Confidence Intervals for Annual Excess Return on the Market

$$Prob(0.08 - 0.2 < R < 0.08 + 0.2) = 68\%$$

$$Prob(0.08 - 2 \times 0.2 < R < 0.08 + 2 \times 0.2) = 95\%$$

$$Prob(0.08 - 3 \times 0.2 < R < 0.08 + 3 \times 0.2) = 99\%$$

The probability that the realization is negative

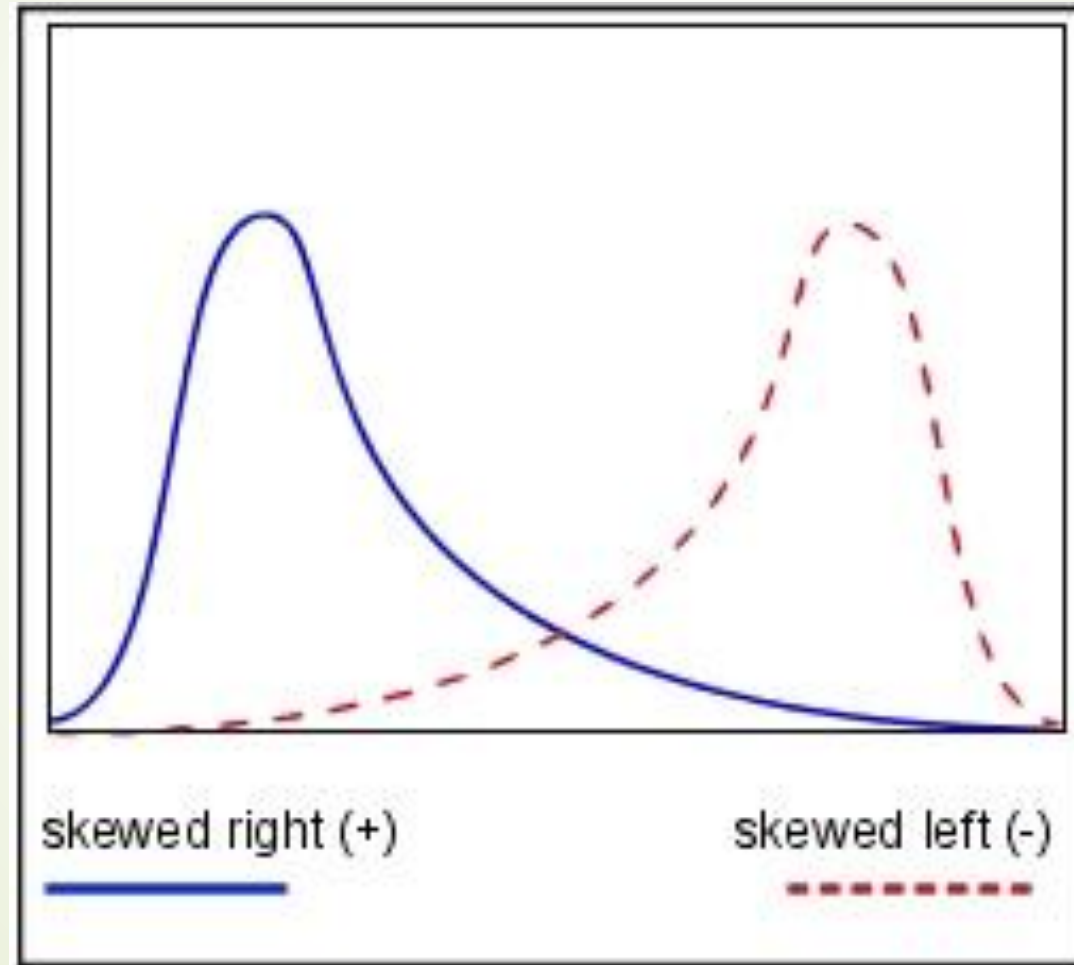
$$Prob(R < 0) = Prob\left(\frac{R-0.08}{0.2} < \frac{0-0.08}{0.2}\right) = Prob(z < -0.4) = 34.4\%$$

# Higher Moments

- Skewness – the third moment is zero.
- Kurtosis – the fourth moment is three times the variance.
- Odd moments are all zero.
- Even moments are (often complex) functions of the mean and the variance.
- In the next slides, skewness and kurtosis are presented for other probability distribution functions.

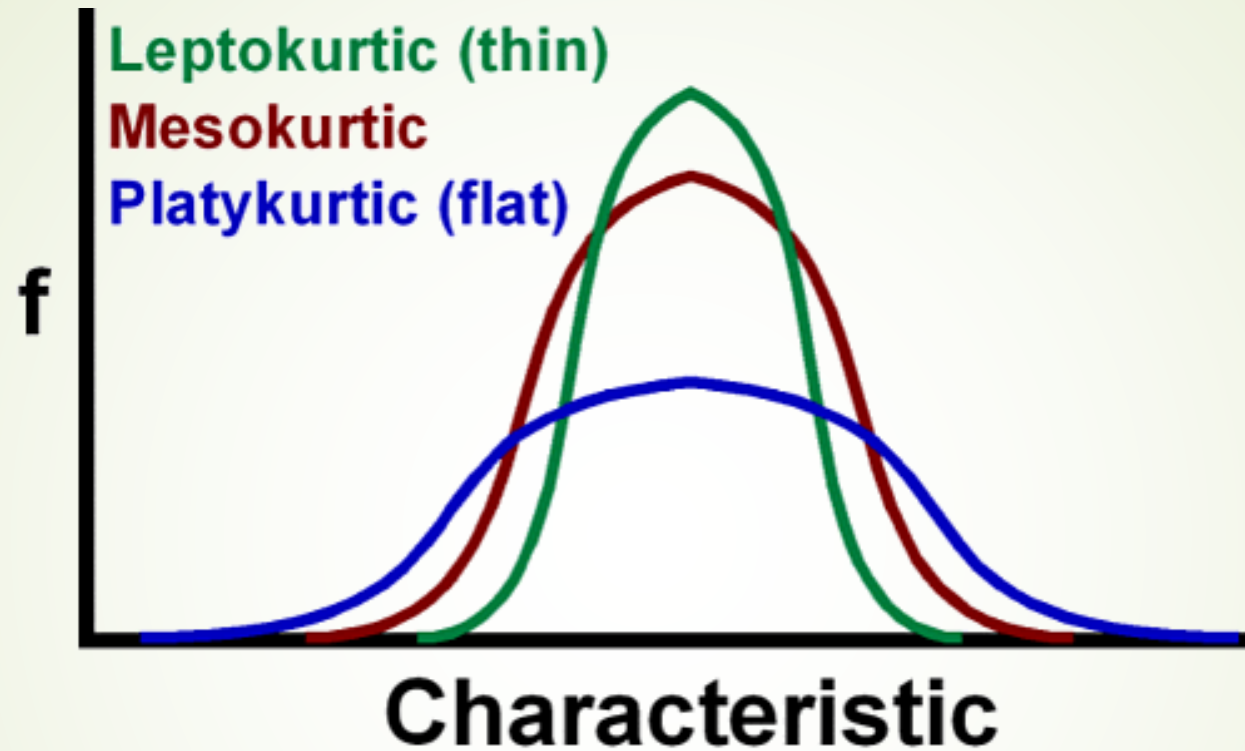
# Skewness

The skewness can be negative (left tail) or positive (right tail).





# Kurtosis



Mesokurtic - A term used in a statistical context where the kurtosis of a distribution is similar, or identical, to the kurtosis of a normally distributed data set.

# The Essence of Skewness and Kurtosis

- Positive skewness means nontrivial probability for large payoffs.
- Kurtosis is a measure for how thick the distribution's tails are. It can reflect the uncertainty about variation.
- When is the skewness zero? In symmetric distributions.

# Bivariate Normal Distribution

➤ Bivariate normal:

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N \left[ \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \right]$$

➤ The marginal densities of  $x$  and  $y$  are

$$x \sim N[(\mu_x, \sigma_x^2)]$$

$$y \sim N[(\mu_y, \sigma_y^2)]$$

➤ What is the distribution of  $y$  if  $x$  is known?

# Conditional Distribution

- The conditional distribution is still normal:

$$y \mid x \sim N \left[ \mu_y + \frac{\sigma_{xy}}{\sigma_x^2} (x - \mu_x), \underbrace{\sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2}}_{\substack{\text{always} \\ \text{non-negative}}} \right]$$

- If the correlation between  $x$  and  $y$  is positive and  $x > \mu_x$  then the conditional expectation is higher than the unconditional one.

# Conditional Moments

- If  $x$  and  $y$  are uncorrelated, then the conditional and unconditional expected return of  $y$  are identical.

- That is, if  $\sigma_{xy} = 0$ , then 
$$\sigma_{y|x} = \sqrt{\sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2}} = \sigma_y$$

meaning that the realization of  $x$  does not say anything about the conditional distribution of  $y$ .

- It should be noted that random variables can be uncorrelated yet dependent.
- Dependence can be represented by a copula.
- Under normality, zero correlation and independence coincide.

# Conditional Standard Deviation

- Developing the conditional standard deviation further:

$$\sigma_{y|x} = \sqrt{\sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2}} = \sqrt{\sigma_y^2 \left(1 - \frac{\sigma_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2}\right)} = \sigma_y \sqrt{(1 - R^2)}$$

- When goodness of fit is higher the conditional standard deviation is lower.
- That makes a great sense: the realization of  $x$  gives substantial information about  $y$ .

# Multivariate Normal

$$X_{m \times 1} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_m \end{bmatrix} \sim N(\mu_{m \times 1}, \Sigma_{m \times m})$$

Define  $Z=AX+B$ .



# Multivariate Normal

$$Z = A_{n \times m} \cdot X_{m \times 1} + B_{n \times 1} \sim N(A_{n \times m} \mu_{m \times 1} + B_{n \times 1}, A_{n \times m} \Sigma_{m \times m} A'_{m \times n})$$

Let us now make some transformations to end up with  $N(0, I)$ :

$$Z - (A_{n \times m} \mu_{m \times 1} + B_{n \times 1}) \sim N(0, A_{n \times m} \Sigma_{m \times m} A'_{m \times n})$$

$$(A_{n \times m} \Sigma_{m \times m} A'_{m \times n})^{-\frac{1}{2}} (Z - A_{n \times m} \mu_{m \times 1} - B_{n \times 1}) \sim N(0, I)$$

# Multivariate Normal

- Consider an  $N$ -vector of stock returns which are normally distributed:

$$R_{N \times 1} \sim N(\mu_{N \times 1}, \Sigma_{N \times N})$$

- Then:

$$R - \mu \sim N(0, \Sigma)$$
$$\Sigma^{-\frac{1}{2}}(R - \mu) \sim N(0, I)$$

- What does  $\Sigma^{-\frac{1}{2}}$  mean? A few rules:

$$\Sigma^{-\frac{1}{2}} \cdot \Sigma^{-\frac{1}{2}} = \Sigma^{-1}$$

$$\Sigma^{\frac{1}{2}} \cdot \Sigma^{\frac{1}{2}} = \Sigma$$

$$\Sigma^{-\frac{1}{2}} \cdot \Sigma^{\frac{1}{2}} = I$$

# The Chi-Squared Distribution

- ▶ If  $X_1 \sim N(0,1)$  then  $X_1^2 \sim \chi^2(1)$
- ▶ If  $X_1 \sim N(0,1)$ ,  $X_2 \sim N(0,1)$ , and  $X_1 \perp X_2$  then

$$X_1^2 + X_2^2 \sim \chi^2(2)$$

# More about the Chi-Squared

Moreover if 
$$\begin{cases} X_1 \sim \chi^2(m) \\ X_2 \sim \chi^2(n) \\ X_1 \perp X_2 \end{cases}$$

then  $X_1 + X_2 \sim \chi^2(m + n)$

# The F Distribution

Gibbons, Ross, and Shanken (GRS) designated a finite sample asset pricing test that has the F - Distribution.

The GRS test is one of the most well known and heavily used in the field of asset pricing.

To understand the F distribution notice that

$$\text{If } \begin{cases} X_1 \sim \chi^2(m) \\ X_2 \sim \chi^2(n) \\ X_1 \perp X_2 \end{cases}$$

$$\text{then } A = \frac{X_1/m}{X_2/n} \sim F(m, n)$$

# The t Distribution

► Suppose that  $r_1, r_2, \dots, r_T$  is a sample of length  $T$  of stock returns which are normally distributed.

► The sample mean and variance are

$$\bar{r} = \frac{r_1 + \dots + r_T}{T} \quad \text{and} \quad s^2 = \frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{r})^2$$

► The following statistic has the  $t$  distribution with  $T-1$  d.o.f

$$t = \frac{\bar{r} - \mu}{s / \sqrt{T-1}}$$

# The t Distribution

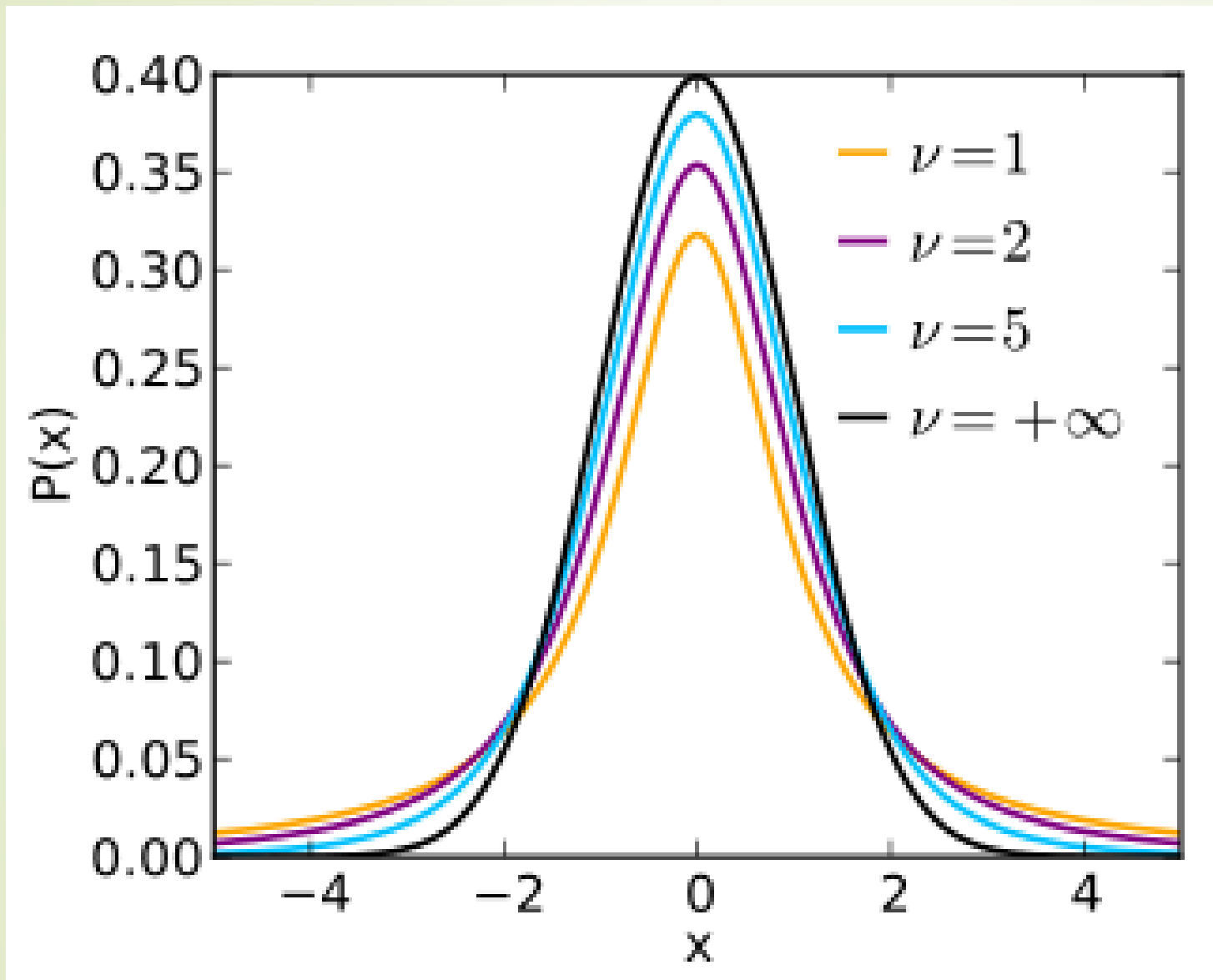
► The *pdf* of student-t is given by:

$$f(x, v) = \frac{1}{\sqrt{v} \cdot B(v/2, 1/2)} \cdot \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}}$$

where  $B(a, b)$  is beta function and  $v \geq 1$  is the number of degrees of freedom



# The Student's t Distribution



$\nu$  is the number of degrees of freedom. When  $\nu = +\infty$  the t-distribution becomes normal dist.

# The $t$ Distribution

- The  $t$ -distribution is the sampling distribution of the  $t$ -value when the sample consist of independently and identically distributed observations from a normally distributed population.
- It is obtained by dividing a normally distributed random variable by a square root of a Chi-squared distributed random variable when both random variables are independent.
- Indeed, later we will show that when returns are normally distributed the sample mean and variance are independent.

# Regression Analysis

- Various applications in corporate finance and asset pricing require the implementation of a regression analysis.
- We will estimate regressions using matrix notation.
- For instance, consider the time series predictive regression

$$R_t = \alpha + \beta_1 Z_{1,t-1} + \beta_2 Z_{2,t-1} + \varepsilon_t \quad t = 1, 2, \dots, T$$

# Rewriting the system in a matrix form

$$R_{T \times 1} = \begin{bmatrix} R_1 \\ R_2 \\ \cdot \\ \cdot \\ R_T \end{bmatrix} \quad X_{T \times 3} = \begin{bmatrix} 1, Z_{1,0}, Z_{2,0} \\ 1, Z_{1,1}, Z_{2,1} \\ \cdot \\ \cdot \\ 1, Z_{1,T-1}, Z_{2,T-1} \end{bmatrix} \quad \gamma_{3 \times 1} = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \varepsilon_{T \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_T \end{bmatrix}$$

yields

$$R = X\gamma + \varepsilon$$

We will derive regression coefficients and their standard errors using OLS, MLE, and Method of Moments.

# Vectors and Matrices: some Rules

➤ A is a column vector:  $A_{t \times 1} = \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \cdot \\ \cdot \\ \alpha_{1t} \end{bmatrix}$

➤ The transpose of A is a row vector:  $A'_{1 \times t} = [\alpha_{11}, \alpha_{12}, \dots, \alpha_{1t}]$

➤ The identity matrix satisfies

$$I \cdot B = B \cdot I = B$$

# Multiplication of Matrices

$$A_{2 \times 2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad B_{2 \times 2} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \quad A'_{2 \times 2} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$$

$$A_{2 \times 2} B_{2 \times 2} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}, \quad (AB)' = B'A' \\ (AB)^{-1} = B^{-1}A^{-1} *$$

The inverse of the matrix  $A$  is  $A^{-1}$  which satisfies:

$$A_{2 \times 2} A_{2 \times 2}^{-1} = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

\* Both  $A$  and  $B$  are invertible

# Solving Large Scale Linear Equations

$$A_{m \times m} X_{m \times 1} = b_{m \times 1}$$
$$X = A^{-1} b$$

Of course,  $A$  has to be a square invertible matrix. That is, the number of equations must be equal to the number of unknowns and further all equations must be linearly independent.



# Linear Independence and Norm

- The vectors  $V_1, \dots, V_N$  are linearly independent if there does not exist scalars  $c_1, \dots, c_N$  such that

$$c_1V_1 + c_2V_2 + \dots + c_NV_N = 0$$

unless  $c_1 = c_2 = \dots = c_N = 0$ .

- In the context of financial economics – we will consider  $N$  risky assets such that the payoff of each asset is not a linear combination of the other  $N-1$  assets.
- Then the covariance matrix of asset returns is positive definite (see next page) of rank  $N$  and hence is invertible.
- The norm of a vector  $V$  is

$$\|V\| = \sqrt{V'V}$$

# A Positive Definite Matrix

- An  $N \times N$  matrix  $\Sigma$  is called positive definite if

$$\begin{matrix} V' & \cdot & \Sigma & \cdot & V & > & 0 \\ 1 \times N & & N \times N & & N \times 1 & & \end{matrix}$$

for any nonzero vector  $V$ .

- Such matrix is invertible and it has  $N$  distinct Eigen-values and Eigen-vectors.
- A non-positive definite matrix cannot be inverted and its determinant is zero.

# The Trace of a Matrix

➔ Let  $\Sigma_{N \times N} = \begin{bmatrix} \sigma_1^2 & \dots & \dots & \dots & \dots & \sigma_{1,N} \\ \sigma_{1,2}, \sigma_2^2 & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_{N,1}, \dots & \dots & \dots & \dots & \dots & \sigma_N^2 \end{bmatrix} = \begin{bmatrix} \sigma^{(1)} \\ \dots \\ \sigma^{(N)} \end{bmatrix}_{N \times 1}$

➔ Trace of a matrix is the sum of diagonal elements

$$tr(\Sigma) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_N^2$$

# Matrix Vectorization

$$\underset{N^2 \times 1}{Vec}(\Sigma) = \begin{bmatrix} \sigma^{(1)} \\ \sigma^{(2)} \\ \vdots \\ \sigma^{(N)} \end{bmatrix}$$

This is the vectorization operator – it works for both square as well as non square matrices.

# The VECH Operator

Similar to  $Vec(\Sigma)$  but takes only the upper triangular elements of  $\Sigma$ .

$$\underset{\frac{(N+1)N}{2} \times 1}{Vech}(\Sigma) = \begin{bmatrix} \sigma_1^{(1)} \\ \sigma_2^2 \\ \vdots \\ \sigma_{2N} \\ \vdots \\ \sigma_N^2 \end{bmatrix}$$

# Partitioned Matrices

➤ A, a square nonsingular matrix, is partitioned as

$$A = \begin{bmatrix} A_{11} & , & A_{12} \\ m_1 \times m_1 & & m_1 \times m_2 \\ A_{21} & , & A_{22} \\ m_2 \times m_1 & & m_2 \times m_2 \end{bmatrix}$$

➤ Then the inverse of A is given by

$$A^{-1} = \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & , & -(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & , & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix}$$

➤ Confirm that  $A \cdot A^{-1} = I_{(m_1+m_2)}$

# Matrix Differentiation

$Y$  is an  $M$ -vector.  $X$  is an  $N$ -vector. Then

$$\frac{\partial Y}{\partial X}_{M \times N} = \begin{bmatrix} \frac{\partial Y_1}{\partial X_1} & \frac{\partial Y_1}{\partial X_2} & \cdots & \frac{\partial Y_1}{\partial X_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial Y_M}{\partial X_1} & \frac{\partial Y_M}{\partial X_2} & \cdots & \frac{\partial Y_M}{\partial X_N} \end{bmatrix}$$

And specifically, if

$$Y_{M \times 1} = A_{M \times N} X_{N \times 1}$$

Then:

$$\frac{\partial Y}{\partial X} = A$$



# Matrix Differentiation

$$\text{Let } Z = \begin{matrix} Y' & A & X \\ 1 \times M & M \times N & N \times 1 \end{matrix}$$

$$\frac{\partial Z}{\partial X} = Y' A$$

$$\frac{\partial Z}{\partial Y} = X' A'$$

# Matrix Differentiation

Let

$$\theta = X' C X$$

$1 \times N$     $N \times N$     $N \times 1$

$$\frac{\partial \theta}{\partial X} = X'(C + C')$$

If  $C$  is symmetric, then

$$\frac{\partial \theta}{\partial X} = 2X' C$$

# Kronecker Product

It is given by

$$C_{np \times mg} = A_{n \times m} \otimes B_{p \times g}$$

$$C_{np \times mg} = \begin{bmatrix} a_{11}B & \dots & \dots & \dots & \dots & \dots & a_{1m}B \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1}B & \dots & \dots & \dots & \dots & \dots & a_{nm}B \end{bmatrix}$$

# Kronecker Product

For square matrices  $A$  and  $B$ , the following equalities apply

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

$$|A_{M \times M} \otimes B_{N \times N}| = |A|^M |B|^N$$

$$(A \otimes B)(C \otimes D) = (AC \otimes BD)$$

# Operations on Matrices: More Advanced

$$\text{tr}(A \otimes B) = \text{tr}(A)\text{tr}(B)$$

$$\text{vec}(A + B) = \text{vec}(A) + \text{vec}(B)$$

$$\text{vech}(A + B) = \text{vech}(A) + \text{vech}(B)$$

$$\text{tr}(AB) = \text{vec}(B')' \text{vec}(A)$$

$$E(X'AX) = E[\text{tr}(X'AX)]$$

$$= E[\text{tr}(XX')A] = \text{tr}[E(XX')A] = \mu' A \mu + \text{tr}(\Sigma A)$$

where

$$\mu = E(X) \text{ and } \Sigma = E(X - \mu)(X - \mu)'$$

# Using Matrix Notation

## in a Portfolio Choice Context

The expectation and the variance of a portfolio's rate of return in the presence of three stocks are formulated as

$$\mu_p = \omega_1\mu_1 + \omega_2\mu_2 + \omega_3\mu_3 = \omega'\mu$$

$$\begin{aligned}\sigma_p^2 &= \omega_1^2\sigma_1^2 + \omega_2^2\sigma_2^2 + \omega_3^2\sigma_3^2 \\ &\quad + 2\omega_1\omega_2\sigma_1\sigma_2\rho_{12} + 2\omega_1\omega_3\sigma_1\sigma_3\rho_{13} + 2\omega_2\omega_3\sigma_2\sigma_3\rho_{23} \\ &= \omega'\Sigma\omega\end{aligned}$$

# Using Matrix Notation

## in a Portfolio Choice Context

where

$$\mu_{3 \times 1} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \quad \omega_{3 \times 1} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} \quad \Sigma_{3 \times 3} = \begin{bmatrix} \sigma_1^2, \sigma_{12}, \sigma_{13} \\ \sigma_{21}, \sigma_2^2, \sigma_{23} \\ \sigma_{31}, \sigma_{32}, \sigma_3^2 \end{bmatrix}$$



# The Expectation, Variance, and Covariance of Sum of Random Variables

$$E(ax + by + cz) = aE(x) + bE(y) + cE(z)$$

$$\begin{aligned} \text{Var}(ax + by + cz) &= a^2\text{Var}(x) + b^2\text{Var}(y) + c^2\text{Var}(z) \\ &\quad + 2ab\text{Cov}(x, y) + 2ac\text{Cov}(x, z) + 2bc\text{Cov}(y, z) \end{aligned}$$

$$\begin{aligned} \text{Cov}(ax + by, cz + dw) &= ac\text{Cov}(x, z) + ad\text{Cov}(x, w) \\ &\quad + bc\text{Cov}(y, z) + bd\text{Cov}(y, w) \end{aligned}$$

# Law of Iterated Expectations (LIE)

- The LIE relates the unconditional expectation of a random variable to its conditional expectation via the formulation

$$E[Y] = E_x\{E[Y | X]\}$$

- Paul Samuelson shows the relation between the LIE and the notion of market efficiency – which loosely speaking asserts that the change in asset prices cannot be predicted using current information.

# LIE and Market Efficiency

- Under rational expectations, the time  $t$  security price can be written as the rational expectation of some fundamental value, conditional on information available at time  $t$ :

$$P_t = E[P^* | I_t] = E_t[P^*]$$

- Similarly,  $P_{t+1} = E[P^* | I_{t+1}] = E_{t+1}[P^*]$

- The conditional expectation of the price change is

$$E[P_{t+1} - P_t | I_t] = E_t[E_{t+1}[P^*] - E_t[P^*]] = 0$$

- The quantity does not depend on the information.

# Variance Decomposition (VD)

- $Var[y]$  can be represented as the sum of two components:

$$Var[y] = Var_x[E(y|x)] + E_x[Var(y|x)]$$

- Shiller (1981) documents excess volatility in the equity market. We can use VD to prove it:

- The theoretical stock price:  $P_t^* = \frac{D_{t+1}}{1+r} + \frac{D_{t+2}}{(1+r)^2} \dots$   
based on actual future dividends

- The actual stock price:  $P_t = E[P_t^* | I_t]$

# Variance Decomposition

$$\text{Var}[P_t^*] = \text{Var}[E(P_t^* | I_t)] + \underbrace{E[\text{Var}(P_t^* | I_t)]}_{\text{positive}}$$

$$\text{Var}[P_t^*] = \text{Var}[P_t] + \text{positive}$$

↓

$$\text{Var}[P_t^*] > \text{Var}[P_t]$$

By variance decomposition, the variance of the theoretical price must be higher than that of the actual price.

Data, however, implies the exact opposite.

Either the present value formula (with constant discount factors) is away off, or asset prices are subject to behavioral biases.

# Session 2(part a) - Taylor Approximations in Financial Economics

# TA in Finance

- Several major applications in finance require the use of Taylor series approximation.
- The following table describes three applications.



# TA in Finance: Major Applications

	Expected Utility	Bond Pricing	Option Pricing
<b>First Order</b>	Mean	Duration	Delta
<b>Second Order</b>	Volatility	Convexity	Gamma
<b>Third Order</b>	Skewness		
<b>Fourth Order</b>	Kurtosis		

# Taylor Approximation

► Taylor series is a representation of a function as an infinite sum of terms that are calculated from the values of the function's derivatives at a single point.

► Taylor approximation is written as:

$$f(x) = f(x_0) + \frac{1}{1!} f'(x_0)(x - x_0) + \\ + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \frac{1}{3!} f'''(x_0)(x - x_0)^3 + \dots$$

► It can also be written with  $\sum$  notation:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^n(x_0)}{n!} (x - x_0)^n$$

# Maximizing Expected Utility of Terminal Wealth

- The invested wealth is  $W_T$ , the investment horizon is  $K$  periods.
- The terminal wealth, the wealth in the end of the investment horizon, is

$$W_{T+K} = W_T(1 + R)$$

where

$$(1 + R) = (1 + R_{T+1})(1 + R_{T+2}) \dots (1 + R_{T+K})$$

- Applying “ln” on both sides of the equation:

$$\begin{aligned} \ln(1 + R) &= \ln((1 + R_{T+1})(1 + R_{T+2}) \dots (1 + R_{T+K})) = \\ &= \ln(1 + R_{T+1}) + \ln(1 + R_{T+2}) + \dots + \ln(1 + R_{T+K}) \end{aligned}$$

# Transforming to Log Returns

► Denote the log returns as:

$$r_{T+1} = \ln ( 1 + R_{T+1} )$$

$$r_{T+2} = \ln ( 1 + R_{T+2} )$$

.....

$$r_{T+K} = \ln ( 1 + R_{T+K} )$$

And:

$$r = \ln ( 1 + R )$$

ent horizon is:

$$r_{T+1} + \dots + r_{T+K}$$

R is therefore:

$$\exp ( r )$$

# Power utility as a Function of Cumulative Log Return (CLR)

- Assume that the investor has the power utility function (where  $0 < \gamma < 1$ ):

$$u(W_{T+K}) = \frac{1}{\gamma} W_{T+K}^{\gamma}$$

- Then the utility function of the terminal wealth can be expressed as a function of CLR

$$u(W_{T+K}) = \frac{1}{\gamma} W_{T+k}^{\gamma} = \frac{1}{\gamma} (W_T \exp(r))^{\gamma} = \frac{1}{\gamma} W_T^{\gamma} \exp(\gamma r)$$

# Power utility as a Function of Cumulative Log Return

- Note that  $r$  is stochastic (unknown), since all future returns are unknown
- We can use the Taylor Approximation to express the utility as a function of the moments of CLR.

# The Utility Function Approximation

$$\begin{aligned}u(W_{T+K}) &= \frac{W_T^\gamma}{\gamma} \exp(\mu\gamma) + \frac{W_T^\gamma}{\gamma} \exp(\mu\gamma)(r - \mu)\gamma + \\ &\frac{1}{2} \frac{W_T^\gamma}{\gamma} \exp(\mu\gamma)(r - \mu)^2\gamma^2 + \frac{1}{6} \frac{W_T^\gamma}{\gamma} \exp(\mu\gamma)(r - \mu)^3\gamma^3 + \\ &\frac{1}{24} \frac{W_T^\gamma}{\gamma} \exp(\mu\gamma)(r - \mu)^4\gamma^4 + \dots\end{aligned}$$

And the expected value of  $u(W_{T+K})$  is:

$$\begin{aligned}E[u(W_{T+K})] &\approx \frac{W_T^\gamma}{\gamma} \exp(\mu\gamma) \left[ 1 + \frac{1}{2} \text{Var}(r)\gamma^2 + \right. \\ &\left. \frac{1}{6} \text{Ske}(r)\gamma^3 + \frac{1}{24} \text{Kur}(r)\gamma^4 \right]\end{aligned}$$



# Expected Utility

- ▶ Let us assume that log return is normally distributed:

$$r \sim N(\mu, \sigma^2) \longrightarrow Ske = 0, Kur = 3\sigma^4$$

Then:

$$E[u(W_{T+K})] \approx \frac{W_T^\gamma}{\gamma} \exp(\gamma\mu) \left( 1 + \frac{\gamma^2}{2} \sigma^2 + \frac{\gamma^4}{8} \sigma^4 \right)$$

- ▶ The exact solution under normality is

$$E[u(W_{T+K})] = \frac{W_T^\gamma}{\gamma} \exp\left(\gamma\mu + \frac{\gamma^2}{2} \sigma^2\right)$$

- ▶ Are approximated and exact solutions close enough?

# Taylor Approximation – Bond Pricing

- Taylor approximation is also used in bond pricing.
- The bond price is:  $P(y_0) = \sum_{i=1}^n \frac{CF_i}{(1+y_0)^i}$  where  $y_0$  is the yield to maturity.
- Assume that  $y_0$  changes to  $y_1$
- The delta (change) of the yield to maturity is written as:  
$$\Delta y = y_1 - y_0$$

# Changes in Yields and Bond Pricing

➤ Using Taylor approximation we get:

$$P(y_1) \approx P(y_0) + \frac{1}{1!} P'(y_0)(y_1 - y_0) + \frac{1}{2!} P''(y_0)(y_1 - y_0)^2$$

➤ Therefore:

$$\begin{aligned} P(y_1) - P(y_0) &\approx \frac{1}{1!} P'(y_0)(y_1 - y_0) + \frac{1}{2!} P''(y_0)(y_1 - y_0)^2 \approx \\ &\approx P'(y_0)\Delta y + \frac{1}{2} P''(y_0) \cdot (\Delta y)^2 \end{aligned}$$

# Duration and Convexity

➔ Denote  $P = P(y_0)$ . Dividing by  $P(y_0)$  yields

$$\frac{\Delta P}{P} \approx \frac{P'(y_0) \cdot \Delta y}{P(y_0)} + \frac{P''(y_0) \cdot (\Delta y)^2}{2P(y_0)}$$

➔ Instead of:  $\frac{P'(y_0)}{P(y_0)}$  we can write “-MD” (Modified Duration).

➔ Instead of:  $\frac{P''(y_0)}{P(y_0)}$  we can write “Con” (Convexity).

# The Approximated Bond Price Change

➔ It is given by

$$\frac{\Delta P}{P} \approx -MD \cdot \Delta y + \frac{Con \cdot \Delta y^2}{2}$$

➔ What if the yield to maturity falls?

# The Bond Price Change when Yields Fall

➤ The change of the bond price is:

$$\frac{\Delta P}{P} = -MD \cdot \Delta y + \frac{Con \cdot \Delta y^2}{2}$$

- According to duration - bond price should increase.
- According to convexity - bond price should also increase.
- The bond price clearly rises.

# The Bond Price Change when Yields Increase

➤ And what if the yield to maturity increases?

➤ Again, the change of the bond price is

$$\frac{\Delta P}{P} = -MD \cdot \Delta y + \frac{Con \cdot \Delta y^2}{2}$$

➤ According to duration – the bond price should decrease.



# The Bond Price Change when Yields Increase

- According to convexity – the bond price should increase.
- What is the overall effect?
- The influence of duration is always stronger than that of convexity as the duration is a first order effect while the convexity is only second order.
- So the bond price must fall.

# Option Pricing

- A call price is a function of the underlying asset price.
- What is the change in the call price when the underlying asset pricing changes?

$$\begin{aligned}C(P_s) &\approx C(P_t) + \frac{\partial C}{\partial P} (P_s - P_t) + \frac{1}{2} \frac{\partial^2 C}{\partial P^2} (P_s - P_t)^2 \\ &\approx C(P_t) + \Delta(P_s - P_t) + \frac{1}{2} \Gamma(P_s - P_t)^2\end{aligned}$$

- Focusing on the first order term – this establishes the delta neutral trading strategy.

# Delta Neutral Strategy

Suppose that the underlying asset volatility increases.

Suppose further that the implied volatility lags behind.

The call option is then underpriced – buy the call.

However, you take the risk of fluctuations in the price of the underlying asset.

To hedge that risk you sell Delta units of the underlying asset.

The same applies to trading strategies involving put options.

# Session 2(part b) - OLS, MLE, MOM

# Ordinary Least Squares (OLS)

- The goal is to estimate the regression parameters (least squares)
- Consider the regression  $y_t = x_t' \beta + \varepsilon_t$  and assume homoskedasticity:

$$Y = \begin{bmatrix} y_1 \\ \cdot \\ y_T \end{bmatrix}, \quad X = \begin{bmatrix} 1, x_{11}, \dots, x_{K1} \\ \dots \dots \dots \dots \dots \\ 1, x_{1T}, \dots, x_{KT} \end{bmatrix}, \quad V = \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \varepsilon_T \end{bmatrix}$$

- Expressing in a matrix form:

$$Y = X\beta + V$$

- Or:

$$V = Y - X\beta$$

- Define a function of the squares of the errors that we want to minimize:

$$f(\beta) = \sum_{t=1}^T \varepsilon_t^2 = V'V = (Y - X\beta)'(Y - X\beta) = Y'Y + \beta'X'X\beta - 2\beta'X'Y$$

# OLS First Order Conditions

- ▶ Let us differentiate the function with respect to beta and consider the first order condition:

$$\frac{\partial f(\beta)}{\partial \beta} = 2(X'X)\beta - 2X'Y = 0$$

$$(X'X)\beta = X'Y$$

$$\Rightarrow \hat{\beta} = (X'X)^{-1}X'Y$$

- ▶ Recall:  $X$  and  $Y$  are observations.
- ▶ Could we choose other  $\beta$  to obtain smaller  $V'V = \sum_{t=1}^T \varepsilon_t^2$  ?

# The OLS Estimator

- No as we minimize the quantity  $V'V$ .
- We know:  $Y = X\beta + V \rightarrow \hat{\beta} = (X'X)^{-1}X'(X\beta + V)$   
$$\hat{\beta} = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'V$$
$$\hat{\beta} = \beta + (X'X)^{-1}X'V$$
- Is the estimator  $\hat{\beta}$  unbiased for  $\beta$  ?

$$\begin{aligned} E[\hat{\beta} - \beta] &= E[(X'X)^{-1}X'V] \\ &= (X'X)^{-1}X'E[V] = 0, \text{ So - it is indeed unbiased.} \end{aligned}$$



# The Standard Errors of the OLS Estimates

► What about the Standard Error of  $\beta$ ?

$$\hat{\Sigma}_{\beta} = E \left[ (\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right]$$

► Reminder:

$$\hat{\beta} - \beta = (X'X)^{-1}X'V$$

$$(\hat{\beta} - \beta)' = V'X(X'X)^{-1}$$

# The Standard Errors of the OLS Estimates

Continuing:

$$\widehat{\Sigma}_{\beta} = E[(X'X)^{-1}X'VV'X(X'X)^{-1}] =$$

$$(X'X)^{-1}X' \overbrace{E[VV']}^{\text{A scalar}} X(X'X)^{-1} = (X'X)^{-1}X'X(X'X)^{-1}\hat{\sigma}_{\varepsilon}^2 = (X'X)^{-1}\hat{\sigma}_{\varepsilon}^2$$

We get:  $\widehat{\Sigma}_{\beta} = (X'X)^{-1}\hat{\sigma}_{\varepsilon}^2$  where  $\hat{\sigma}_{\varepsilon}^2 = \frac{1}{T-K-1}\widehat{V}'\widehat{V}$  and  $K$  is the number of explanatory variables.

Therefore, we can calculate  $\widehat{\Sigma}_{\beta}$ :

$$\widehat{\Sigma}_{\beta} = \frac{(X'X)^{-1}\widehat{V}'\widehat{V}}{T-K-1}$$

# Maximum Likelihood Estimation (MLE)

- We now turn to Maximum Likelihood as a tool for estimating parameters as well as testing models.
- Assume that  $r_t \stackrel{iid}{\sim} N(\mu, \sigma^2)$
- The goal is to estimate the distribution of the underlying parameters:

$$\theta = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}.$$

Intuition: the data implies the “most likely” value of  $\theta$ .

- MLE is an asymptotic procedure and it is a parametric approach in that the distribution of the regression residuals must be specified explicitly.

# Implementing MLE

- ▶ Let us estimate  $\mu$  and  $\sigma^2$  using MLE; then derive the joint distribution of these estimates.
- ▶ Under normality, the probability distribution function (pdf) of the rate of return takes the form

$$pdf(r_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \frac{(r_t - \mu)^2}{\sigma^2} \right]$$

# The Joint Likelihood

► We define the “Likelihood function”  $L$  as the joint pdf

► Following Bayes Rule:

$$L = pdf(r_1, r_2, \dots, r_T) = pdf(r_1 | r_2 \dots r_T) \times pdf(r_2 | r_3 \dots r_T) \times \dots \times pdf(r_T)$$

► Since returns are assumed *IID* - it follows that

$$L = pdf(r_1) \times pdf(r_2) \times \dots \times pdf(r_T)$$

$$L = (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left[ -\frac{1}{2} \sum_{t=1}^T \left( \frac{r_t - \mu}{\sigma} \right)^2 \right]$$

► Now take the natural log of the joint likelihood:

$$\ln L = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2} \sum_{t=1}^T \left( \frac{r_t - \mu}{\sigma} \right)^2$$

# MLE: Sample Estimates

- Derive the first order conditions

$$\frac{\partial \ln L}{\partial \mu} = \sum_{t=1}^T \frac{r_t - \mu}{\sigma^2} = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2} \sum_{t=1}^T \frac{(r_t - \mu)^2}{\sigma^4} = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (r_t - \hat{\mu})^2$$

- Since  $E[\hat{\sigma}^2] \neq \sigma^2$  - the variance estimator is not unbiased.

# MLE: The Information Matrix

► Take second derivatives:

$$\frac{\partial^2 \ln L}{\partial \mu^2} = - \sum_{t=1}^T \frac{1}{\sigma^2} = - \frac{T}{\sigma^2} \quad \Rightarrow \quad E \left( \frac{\partial^2 \ln L}{\partial \mu^2} \right) = - \frac{T}{\sigma^2}$$

$$\frac{\partial^2 \ln L}{\partial \mu \partial \sigma^2} = - \sum_{t=1}^T \frac{r_t - \mu}{\sigma^4} \quad \Rightarrow \quad E \left( \frac{\partial^2 \ln L}{\partial \mu \partial \sigma^2} \right) = 0$$

$$\frac{\partial^2 \ln L}{(\partial \sigma^2)^2} = \frac{T}{2\sigma^4} - \sum_{t=1}^T \frac{(r_t - \mu)^2}{\sigma^4} \quad \Rightarrow \quad E \left( \frac{\partial^2 \ln L}{(\partial \sigma^2)^2} \right) = - \frac{T}{2\sigma^4}$$



# MLE: The Covariance Matrix

► Set the information matrix  $I(\theta) = -E \left( \frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \right)$

► The variance of  $\hat{\theta}$  is:

$$\text{Var}(\hat{\theta}) = I(\theta)^{-1}$$

Or put differently, the asymptotic distribution of  $\hat{\theta}$  is:

$$(\theta - \hat{\theta}) \sim N(0, I(\theta)^{-1})$$

[It is suggested that you find and understand the proof]

► In our context, the covariance matrix is derived from the information matrix as follows:

$$\begin{bmatrix} -\frac{T}{\sigma^2}, 0 \\ 0, -\frac{T}{2\sigma^4} \end{bmatrix} \xrightarrow{\text{MULTIPLY "-1"}} \begin{bmatrix} \frac{T}{\sigma^2}, 0 \\ 0, \frac{T}{2\sigma^4} \end{bmatrix} \xrightarrow{\text{INVERSE}} \begin{bmatrix} \frac{\sigma^2}{T}, 0 \\ 0, \frac{2\sigma^4}{T} \end{bmatrix}$$

# To Summarize

► Multiply by  $\sqrt{T}$  and get:

$$\sqrt{T}(\theta - \hat{\theta}) \sim N(0, T \cdot I(\theta)^{-1})$$

► And more explicitly:

$$\sqrt{T} \begin{bmatrix} \mu - \frac{1}{T} \sum_{t=1}^T r_t \\ \sigma^2 - \frac{1}{T} \sum_{t=1}^T (r_t - \frac{1}{T} \sum_{t=1}^T r_t)^2 \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2, 0 \\ 0, 2\sigma^4 \end{bmatrix} \right)$$

# The Sample Mean and Variance

- Notice that when returns are normally distributed – the sample mean and the sample variance are independent.
- Departing from normality, the covariance between the sample mean and variance is nonzero.
- Notice also that the ratio obtained by dividing the variance of the variance ( $2\sigma^4$ ) by the variance of the mean ( $\sigma^2$ ) is smaller than one as long as volatility is below 70%.
- The mean return estimate is more noisy because the volatility is typically far below 70%, especially for well diversified portfolios.

# Method of Moments: departing from Normality

➤ We know:

$$E(r_t) = \mu$$

$$E(r_t - \mu)^2 = \sigma^2$$

➤ If:

$$g_t(\theta) = \begin{bmatrix} r_t - \mu \\ (r_t - \mu)^2 - \sigma^2 \end{bmatrix}$$

➤ Then:

$$E(g_t) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

➤ That is, we set two momentum conditions.

# Method of Moments (MOM)

➤ There are two parameters:  $\mu, \sigma^2$

➤ Stage 1: Moment Conditions  $g_t = \begin{bmatrix} r_t - \mu \\ (r_t - \mu)^2 - \sigma^2 \end{bmatrix}$

➤ Stage 2: estimation

$$\frac{1}{T} \sum_{t=1}^T \hat{g}_t = 0$$

# Method of Moments

Continue estimation:

$$\frac{1}{T} (r_t - \hat{\mu}) = 0 \quad t = 1, \dots, T$$

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t$$

$$\frac{1}{T} \sum_{t=1}^T [(r_t - \hat{\mu})^2 - \hat{\sigma}^2] = 0$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (r_t - \hat{\mu})^2$$

# MOM: Stage 3

$$S = E(g_t g_t')$$

$$S = E \begin{bmatrix} (r_t - \mu)^2, & (r_t - \mu)^3 - \sigma^2(r_t - \mu) \\ (r_t - \mu)^3 - \sigma^2(r_t - \mu), & (r_t - \mu)^4 - 2\sigma^2(r_t - \mu)^2 + \sigma^4 \end{bmatrix}$$

↓

$$S = \begin{bmatrix} \sigma^2, & \mu_3 \\ \mu_3, & \mu_4 - \sigma^4 \end{bmatrix}$$



# MOM: stage 4

➔ Memo:

$$g_t(\theta) = \begin{bmatrix} r_t - \mu \\ (r_t - \mu)^2 - \sigma^2 \end{bmatrix}$$

➔ Stage 4: differentiate  $g_t(\theta)$  w.r.t.  $\theta$  and take the expected value

$$D = E \left[ \frac{\partial g_t(\theta)}{\partial \theta} \right] = E \begin{bmatrix} -1, & 0 \\ -2(r_t - \mu), & -1 \end{bmatrix} = \begin{bmatrix} -1, & 0 \\ 0, & -1 \end{bmatrix}$$

# MOM: The Covariance Matrix

Stage 5:

$$\Sigma_{\theta} = (D'S^{-1}D)^{-1}$$

In this specific case we have:  $D = -I$ , therefore:

$$\Sigma_{\theta} = S$$

# The Covariance Matrices

Denote the MLE covariance matrix by  $\Sigma_{\theta}^1$ , and the MOM covariance matrix by  $\Sigma_{\theta}^2$ :

$$\Sigma_{\theta}^1 = \begin{bmatrix} \sigma^2, & 0 \\ 0, & 2\sigma^4 \end{bmatrix}, \quad \Sigma_{\theta}^2 = \begin{bmatrix} \sigma^2, & \mu_3 \\ \mu_3, & \mu_4 - \sigma^4 \end{bmatrix}$$

$$\mu_3 = SK = sk \times \sigma^3$$

(sk is the skewness of the standardized return)

$$\mu_4 = KR = kr \times \sigma^4$$

(kr is the kurtosis of the standardized return)

► Sample estimates of the Skewness and Kurtosis are

$$\hat{\mu}_3 = \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^3 \quad \hat{\mu}_4 = \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^4$$

# Under Normality the MOM Covariance Matrix Boils Down to MLE

$$\Sigma_{\theta}^2 = \begin{bmatrix} \sigma^2, & \mu_3 = 0 \\ \mu_3 = 0, & \mu_4 - \sigma^4 = 2\sigma^4 \end{bmatrix} = \Sigma_{\theta}^1$$

# MOM: Estimating Regression Parameters

- ▶ Let us run the time series regression

$$r_t = \alpha + \beta \cdot r_{mt} + \varepsilon_t$$

where:

$$x_t = [1, r_{mt}]'$$

$$\theta = [\alpha, \beta]'$$

- ▶ We know that  $E[\varepsilon_t | x_t] = 0$
- ▶ Given that  $E[\varepsilon_t | x_t] = 0$ , from the Law of Iterated Expectations (LIE) it follows that  $E[x_t \varepsilon_t] = [x_t (r_t - x_t' \theta)] = 0$
- ▶ That is, there are two moment conditions (stage 1):

$$E(\varepsilon_t) = 0$$

$$E(r_{mt} \varepsilon_t) = 0$$

→

$$g_t = \begin{bmatrix} r_t - \alpha - \beta \cdot r_{mt} \\ (r_t - \alpha - \beta \cdot r_{mt}) r_{mt} \end{bmatrix}$$

# MOM: Estimating Regression Parameters

Stage 2: estimation:

$$\frac{1}{T} \sum_{t=1}^T x_t r_t - \frac{1}{T} \sum_{t=1}^T x_t x_t' \hat{\theta} = 0$$

$$\hat{\theta} = \left( \sum_{t=1}^T x_t x_t' \right)^{-1} \sum_{t=1}^T x_t r_t = (X'X)^{-1} X'R$$

$$X_{T \times 2} = \begin{bmatrix} 1, & r_{m1} \\ \dots & \dots \\ 1, & r_{mT} \end{bmatrix} \quad R = \begin{bmatrix} r_1 \\ \dots \\ r_T \end{bmatrix}$$

# MOM: Estimating Standard Errors

Estimation of the covariance matrix (assuming no serial correlation)

Stage 3:

$$S_T = \frac{1}{T} \sum_{t=1}^T g_t(\hat{\theta}) g_t(\hat{\theta})' = \frac{1}{T} \sum_{t=1}^T (x_t x_t') \hat{\varepsilon}_t^2$$

Stage 4:

$$D_T = \frac{1}{T} \sum_{t=1}^T \frac{\partial g_t(\hat{\theta})}{\partial \hat{\theta}} = -\frac{1}{T} \sum_{t=1}^T (x_t x_t') = -\frac{X'X}{T}$$



# MOM: Estimating Standard Errors

► Stage 5: the covariance matrix estimate is:

$$\Sigma_{\theta} = (D_T' S_T^{-1} D_T)^{-1}$$

$$\Sigma_{\theta} = T(X'X)^{-1} \left[ \sum_{t=1}^T (x_t x_t') \hat{\varepsilon}_t^2 \right] (X'X)^{-1}$$

► Then, asymptotically we get

$$\sqrt{T}(\theta - \hat{\theta}) \sim N(0, \Sigma_{\theta})$$

# Session #3: Hypothesis Testing

# Overview

A short brief of the major contents for today's class:

- Hypothesis testing
- TESTS: Skewness, Kurtosis, Bera-Jarque
- Deriving test statistic for the Sharpe ratio

# Hypothesis Testing

- Let us assume that a mutual fund invests in value stocks (e.g., stocks with high ratios of book-to-market).
- Performance evaluation is mostly about running the regression of excess fund returns on the market benchmark (often multiple benchmarks):

$$R_{it}^e = \alpha_i + \beta_i R_{mt}^e + \varepsilon_{it}$$

- The hypothesis testing for examining performance is
  - $H_0: \alpha_i = 0$  means no performance
  - $H_1$ : Otherwise (positive or negative performance)

# Hypothesis Testing

- Errors emerge if we reject  $H_0$  while it is true, or when we do not reject  $H_0$  when it is wrong:

True state of world



	Don't reject $H_0$	Reject $H_0$
$H_0$	Good decision	Type 1 error $\alpha$
$H_1$	Type 2 error $\beta$	Good decision

# Hypothesis Testing - Errors

- $\alpha$  is the first type error (size), while  $\beta$  is the second type error (related to power).
- The power of the test is equal to  $1 - \beta$ .
- We would prefer both  $\alpha$  and  $\beta$  to be as small as possible, but there is always a trade-off.
- When  $\alpha$  decreases  $\rightarrow \beta$  increases and vice versa.
- The implementation of hypothesis testing requires the knowledge of distribution theory.

# Skewness, Kurtosis & Bera Jarque Test Statistics

- We aim to test normality of stock returns.
- We use three distinct tests to examine normality.



# Test 1 – Skewness

## ➤ TEST 1 - Skewness (third moment)

➤ The setup for testing normality of stock return:

$$H_0: R_t \sim N(\mu, \sigma^2)$$

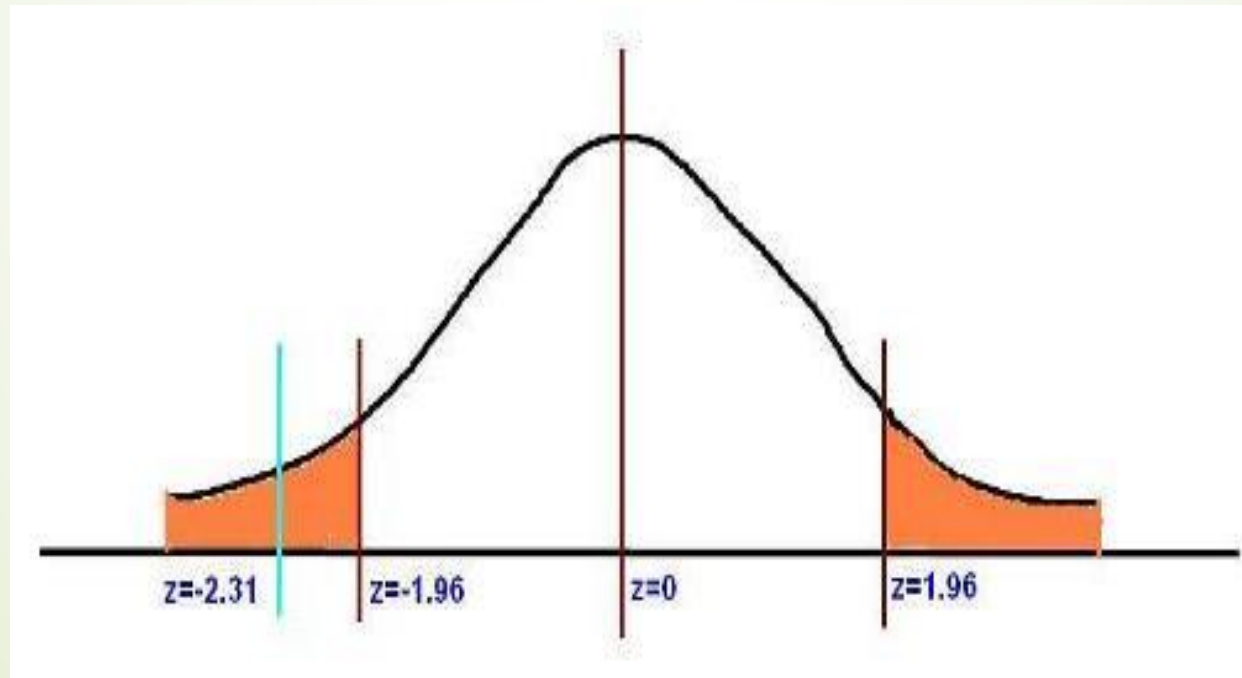
$H_1$ : otherwise

➤ Sample Skewness is

$$S = \frac{1}{T} \sum_{t=1}^T \left( \frac{R_t - \hat{\mu}}{\hat{\sigma}} \right)^3 \stackrel{H_0}{\sim} N \left( 0, \frac{6}{T} \right)$$

# Test I – Skewness

- ▶ Multiplying  $S$  by  $\sqrt{\frac{T}{6}}$ , we get  $\sqrt{\frac{T}{6}}S \sim N(0,1)$
- ▶ If the statistic value is higher (absolute value) than the critical value e.g., the statistic is equal to -2.31, then reject  $H_0$ , otherwise do not reject the null of normality.



# TEST 2 - Kurtosis

➤ Kurtosis estimate is:

$$K = \frac{1}{T} \sum_{t=1}^T \left( \frac{R_t - \hat{\mu}}{\hat{\sigma}} \right)^4 \stackrel{H_0}{\sim} N \left( 3, \frac{24}{T} \right)$$

➤ After transformation:

$$\sqrt{\frac{T}{24}} (K - 3) \stackrel{H_0}{\sim} N(0,1)$$

# TEST 3 - Bera-Jarque Test

➔ The statistic is:

$$BJ = \frac{T}{6} S^2 + \frac{T}{24} (K - 3)^2 \sim \chi^2(2)$$

➔ Why  $\chi^2(2)$  ?

# TEST 3 - Bera-Jarque Test

➔ If  $X_1 \sim N(0,1)$ ,  $X_2 \sim N(0,1)$ , and  $X_1 \perp X_2$  then:

$$X_1^2 + X_2^2 \sim \chi^2(2)$$

➔  $\sqrt{\frac{T}{6}} S$  and  $\sqrt{\frac{T}{24}} (K - 3)$  are both standard normal and they are independent random variables.

# Chi Squared Test

➤ In financial economics, the Chi squared test is implemented quite frequently in hypothesis testing.

➤ Let us derive it.

➤ Suppose that:  $y = \Sigma^{-\frac{1}{2}}(R - \mu) \sim N(0, I)$

Then:  $y'y = (R - \mu)' \Sigma^{-1}(R - \mu) \sim \chi^2(N)$

# Joint Hypothesis Test

You run the regression:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t, \quad t = 1, 2, \dots, T$$



# Joint Hypothesis Test

- GMM would establish three orthogonal conditions:
- $$E(\varepsilon_t) = 0$$
- $$E(\varepsilon_t x_{1t}) = 0$$
- $$E(\varepsilon_t x_{2t}) = 0$$

- Using matrix notation:

$$Y_{T \times 1} = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_T \end{bmatrix} \quad X_{T \times 3} = \begin{bmatrix} 1, x_{11}, x_{21} \\ 1, x_{12}, x_{22} \\ \dots \dots \dots \\ 1, x_{1T}, x_{2T} \end{bmatrix} \quad E_{T \times 1} = \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \varepsilon_T \end{bmatrix} \quad \beta_{3 \times 1} = [\beta_1, \beta_2, \beta_3]'$$

$$Y_{T \times 1} = X_{T \times 3} \beta_{3 \times 1} + E_{T \times 1}$$

# Joint Hypothesis Test

► Let us assume that:  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$      $\varepsilon_1, \varepsilon_2, \varepsilon_3 \dots \varepsilon_T$  are *iid*

↓

$$\hat{\beta} \sim N(\beta, (X'X)^{-1}\sigma_\varepsilon^2)$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

► Joint hypothesis testing:

$$H_0: \beta_0 = 1, \beta_2 = 0$$

$$H_1: \textit{otherwise}$$

# Joint Hypothesis Test

➔ Define:  $R = \begin{bmatrix} 1,0,0 \\ 0,0,1 \end{bmatrix}$ ,  $q = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

➔ The joint test becomes:

$$H_0: R\beta = q$$
$$R \rightarrow \begin{bmatrix} 1,0,0 \\ 0,0,1 \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_2 \end{bmatrix} \stackrel{H_0}{=} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \leftarrow q$$
$$H_1: \textit{otherwise}$$

# Joint Hypothesis Test

- Returning to the testing:  $H_0: R\beta - q = 0$   
 $H_1: \textit{otherwise}$

↓

$$R\hat{\beta} - q \sim N(R\beta - q, R\Sigma_{\beta}R')$$

- Under  $H_0$ :  $R\hat{\beta} - q \stackrel{H_0}{\sim} N(0, R\Sigma_{\beta}R')$

- Chi squared:  $\underbrace{(R\hat{\beta} - q)'(R\Sigma_{\beta}R')^{-1}(R\hat{\beta} - q)}_{\textit{test statistic}} \sim \chi^2(2)$

# Joint Hypothesis Test

Yet, another example:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + \varepsilon_t \\ t = 1, 2, 3, \dots, T$$

$$Y_{T \times 1} = \begin{bmatrix} y_1 \\ \cdot \\ y_T \end{bmatrix}, \quad X_{T \times 6} = \begin{bmatrix} 1, x_{11}, \dots, x_{51} \\ \cdot \\ \cdot \\ 1, x_{1T}, \dots, x_{5T} \end{bmatrix}, \quad \beta_{6 \times 1} = [\beta_0, \beta_1, \dots, \beta_6]' \\ \varepsilon_{T \times 1} = \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \varepsilon_T \end{bmatrix}$$

$$Y = X\beta + E$$

# Joint Hypothesis Test

► Joint hypothesis test:

$$H_0: \beta_0 = 1, \beta_1 = \frac{1}{2}, \beta_2 = \frac{1}{3}, \beta_3 = \frac{1}{4}, \beta_5 = 7$$
$$H_1: \textit{otherwise}$$

► Here is a receipt:

1.  $\hat{\beta} = (X'X)^{-1}X'Y$
2.  $\hat{E} = Y - X\hat{\beta}$
3.  $\hat{\sigma}_\varepsilon^2 = \frac{1}{T-6} \hat{E}' \cdot \hat{E}$
4.  $\hat{\beta} \sim N(\beta, (X'X)^{-1}\sigma_\varepsilon^2 = \Sigma_\beta)$

# Joint Hypothesis Test

$$5. \quad R = \begin{bmatrix} 1,0,0,0,0,0 \\ 0,1,0,0,0,0 \\ 0,0,1,0,0,0 \\ 0,0,0,1,0,0 \\ 0,0,0,0,0,1 \end{bmatrix}, q = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \\ 3 \\ 1 \\ 4 \\ 7 \end{bmatrix}$$

$$6. \quad H_0: R\beta - q = 0$$

$$7. \quad R\hat{\beta} - q \sim N(R\beta - q, R \Sigma_{\beta} R')$$

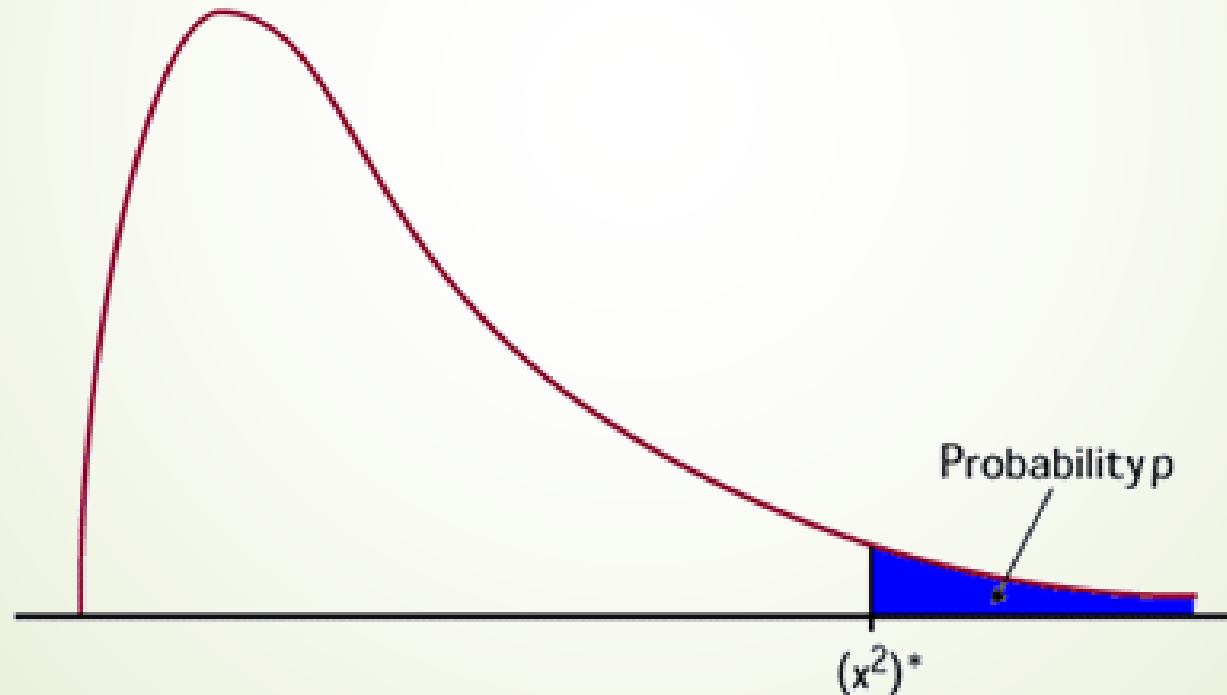
$$8. \quad R\hat{\beta} - q \stackrel{H_0}{\sim} N(0, R \Sigma_{\beta} R')$$

$$9. \quad (R\hat{\beta} - q)' (R \Sigma_{\beta} R')^{-1} (R\hat{\beta} - q) \stackrel{H_0}{\sim} \chi^2(5)$$



# Joint Hypothesis Test

- There are five degrees of freedom implied by the five restrictions on  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_5$
- The chi-squared distribution is always positive.



# Estimating the Sample Sharpe Ratio

- You observe time series of returns on a stock, or a bond, or any investment vehicle (e.g., a mutual fund or a hedge fund):  
 $(r_1, r_2, \dots, r_T)$
- You attempt to estimate the mean and the variance of those returns, derive their distribution, and test whether the Sharpe Ratio of that investment is equal to zero.
- Let us denote the set of parameters by  $\theta = [\mu, \sigma^2]'$
- The Sharpe ratio is equal to  $SR(\theta) = \frac{\mu - r_f}{\sigma}$

# MLE vs. MOM

- To develop a test statistic for the SR, we can implement the MLE or MOM, depending upon our assumption about the return distribution.
- Let us denote the sample estimates by  $\hat{\theta}$


$$r_t \stackrel{iid}{\sim} N(\mu, \sigma^2)$$

**MLE**

$$\sqrt{T}(\hat{\theta} - \theta) \stackrel{a}{\sim} N(0, \Sigma_{\theta}^1)$$

*returns depart from  
iid normal*

**MOM**

$$\sqrt{T}(\hat{\theta} - \theta) \stackrel{a}{\sim} N(0, \Sigma_{\theta}^2)$$

# MLE vs. MOM

As shown earlier, the asymptotic distribution using either MLE or MOM is normal with a zero mean but distinct variance covariance matrices:

$$\sqrt{T}(\hat{\theta} - \theta) \stackrel{MLE}{\sim} N(0, \Sigma_{\theta}^1) \quad \sqrt{T}(\hat{\theta} - \theta) \stackrel{MOM}{\sim} N(0, \Sigma_{\theta}^2)$$

# Distribution of the SR Estimate: The Delta Method

- We will show that  $\sqrt{T}(S\hat{R} - SR) \overset{a}{\sim} N(0, \sigma_{SR}^2)$
- We use the Delta method to derive  $\sigma_{SR}^2$
- The delta method is based upon the first order Taylor approximation.

# Distribution of the SR Estimate: The Delta Method

➔ The first-order TA is

$$SR(\theta) = SR(\hat{\theta}) + \frac{\partial SR}{\partial \theta'}_{1 \times 2} \cdot (\theta - \hat{\theta})_{2 \times 1} \Rightarrow$$

$$SR(\theta) - SR(\hat{\theta}) = \frac{\partial SR}{\partial \theta'}_{1 \times 2} \cdot (\theta - \hat{\theta})_{2 \times 1}$$

➔ The derivative is estimated at  $\hat{\theta}$

# Distribution of the Sample SR

$$E[SR(\theta) - SR(\hat{\theta})] = E \left[ \frac{\partial SR}{\partial \theta'} (\theta - E(\hat{\theta})) \right] = 0$$

whereas

$$VAR(\hat{\theta}) = E [(\hat{\theta} - \theta)(\hat{\theta} - \theta)']$$

$$VAR[SR(\theta) - SR(\hat{\theta})] = E [SR(\theta) - SR(\hat{\theta})]^2$$



# The Variance of the SR

Continue:

$$\begin{aligned} E \left[ \frac{\partial SR}{\partial \theta'} (\theta - \hat{\theta})(\theta - \hat{\theta})' \frac{\partial SR}{\partial \theta} \right] &= \\ &= \frac{\partial SR}{\partial \theta'} \left[ E \left[ (\theta - \hat{\theta})(\theta - \hat{\theta})' \right] \right] \frac{\partial SR}{\partial \theta} = \\ &= \frac{\partial SR}{\partial \theta'} \Sigma_{\theta} \frac{\partial SR}{\partial \theta} \end{aligned}$$

# First Derivatives of the SR

➤ The SR is formulated as  $SR = \frac{\mu - r_f}{(\sigma^2)^{0.5}}$

➤ Let us derive:

$$\frac{\partial SR}{\partial \mu} = \frac{1}{\sigma}$$

$$\frac{\partial SR}{\partial \sigma^2} = \frac{-\frac{1}{2} (\sigma^2)^{-0.5} (\mu - r_f)}{\sigma^2} = \frac{-(\mu - r_f)}{2\sigma^3}$$

# The Distribution of SR under MLE

Continue:

$$\frac{\partial SR}{\partial \theta'} \Sigma_{\theta} \frac{\partial SR}{\partial \theta} = \left( \frac{1}{\sigma}, -\left( \frac{\mu - r_f}{2\sigma^3} \right) \right) \begin{pmatrix} \sigma^2, & 0 \\ 0, & 2\sigma^4 \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma} \\ -\left( \frac{\mu - r_f}{2\sigma^3} \right) \end{pmatrix}$$
$$\Rightarrow \sqrt{T}(SR - S\hat{R}) \sim N \left( 0, \left( 1 + \frac{1}{2} S\hat{R}^2 \right) \right)$$

# The Delta Method in General

➤ Here is the general application of the delta method

➤ If  $\sqrt{T}(\hat{\theta} - \theta) \sim N(0, \Sigma_{\theta})$

➤ Then let  $d(\theta)$  be some function of  $\theta$  :

$$\sqrt{T} \left( d(\hat{\theta}) - d(\theta) \right) \sim N \left( 0, D(\theta) \times \Sigma_{\theta} \times D(\theta)' \right)$$

where  $D(\theta)$  is the vector of derivatives of  $d(\theta)$  with respect to  $\theta$

# Hypothesis Testing

- Does the S&P index outperform the  $R_f$ ?

$$H_0: SR=0$$

$H_1$ : Otherwise

- Under the null there is no outperformance.

$$\sqrt{T}S\hat{R} \sim N(0, 1 + \frac{1}{2}S\hat{R}^2)$$

- Thus, under the null

$$\frac{\sqrt{T}S\hat{R}}{\sqrt{1 + \frac{1}{2}S\hat{R}^2}} \sim N(0, 1)$$

# Session #4: The Efficient Frontier and the Tangency Portfolio

# Testing Asset Pricing Models

- Central to financial econometrics is the formation of test statistics to examine the validity of asset pricing models.
- There are time series as well as cross sectional asset pricing tests.
- In this course the focus is on time-series tests, while in the companion course for PhD students – Asset Pricing – also cross sectional tests are covered.

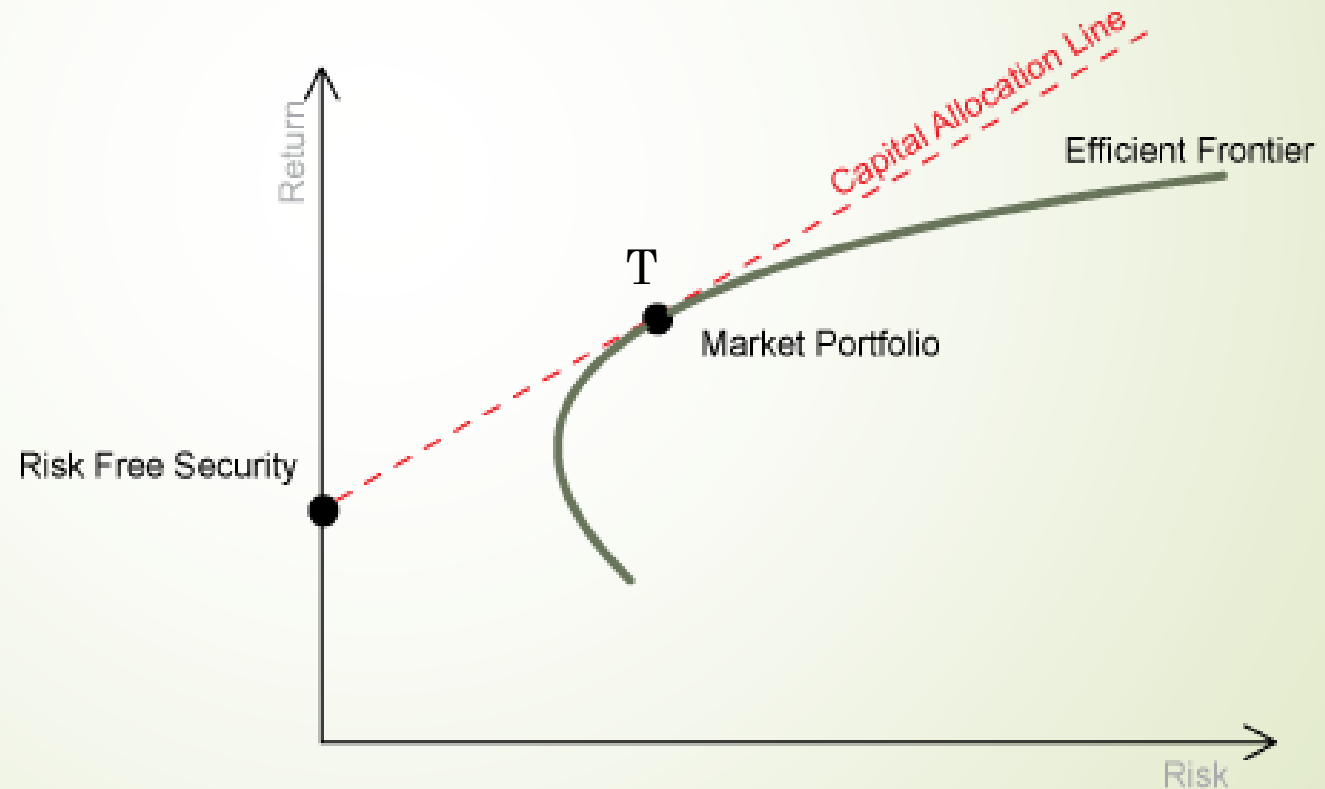


# Testing Asset Pricing Models

- ▶ Time series tests are only implementable when common factors are portfolio spreads, such as excess return on the market portfolio as well as the SMB (small minus big), the HML (high minus low), the WML (winner minus loser), the TERM (long minus short maturity), and the DEF (low minus high quality) portfolios.
- ▶ The first four are equity while the last two are bond portfolios.
- ▶ Cross sectional tests apply to both portfolio and non-portfolio based factors.
- ▶ Consumption growth in the consumption based CAPM (CCAPM) is a good example of a factor that is not a return spread.

# Time Series Tests and the Tangency Portfolio

- Interestingly, time series tests are directly linked to the notion of the tangency portfolio and the efficient frontier.
- Here is the efficient frontier, in which the tangency portfolio is denoted by T.



# Economic Interpretation of the Time Series Tests

- ▶ Testing the validity of the CAPM entails the time series regressions:

$$r_{1t}^e = \alpha_1 + \beta_1 r_{mt}^e + \varepsilon_{1t}$$

.....

$$r_{Nt}^e = \alpha_N + \beta_N r_{mt}^e + \varepsilon_{Nt}$$

- ▶ The CAPM says:  $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_N = 0$   
 $H_1: \text{Otherwise}$

# Economic Interpretation of the Time Series Tests

- The null is equivalent to the hypothesis that the market portfolio is the tangency portfolio.
- Of course, even if the model is valid – the market portfolio **WILL NEVER** lie on the estimated frontier.
- This is due to sampling errors; the efficient frontier is estimated.
- The question is whether the market portfolio is close enough, up to a statistical error, to the tangency portfolio.

# What about Multi-Factor Models?

- The CAPM is a one-factor model.
- There are several extensions to the CAPM.
- The multivariate version is given by the  $K$ -factor model:

# Testing Multifactor Models

$$r_{1t}^e = \alpha_1 + \beta_{11}f_1 + \beta_{12}f_2 + \cdots + \beta_{1K}f_K + \varepsilon_{1t}$$

.....

$$r_{Nt}^e = \alpha_N + \beta_{N1}f_1 + \beta_{N2}f_2 + \cdots + \beta_{NK}f_K + \varepsilon_{Nt}$$

- The null is again:  $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_N = 0$   
 $H_1: \text{Otherwise}$
- In the multi-factor context, the hypothesis is that some particular combination of the factors is the tangency portfolio.

# The Efficient Frontier: Investable Assets

- Consider  $N$  risky assets whose returns at time  $t$  are:

$$R_t = \begin{bmatrix} R_{1t} \\ \dots \\ R_{Nt} \end{bmatrix}$$

$N \times 1$

- The expected value of return is denoted by:

$$E(R_t) = \begin{bmatrix} E(R_{1t}) \\ \dots \\ E(R_{Nt}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \dots \\ \mu_N \end{bmatrix} = \mu$$



# The Covariance Matrix

► The variance covariance matrix is denoted by:

$$V = E[(R_t - \mu)(R_t - \mu)'] = \begin{bmatrix} \sigma_1^2, \dots, \dots, \dots \sigma_{1N} \\ \dots, \sigma_2^2, \dots, \dots \sigma_{2N} \\ \dots \dots \dots \dots \dots \\ \dots \dots \dots \dots, \sigma_N^2 \end{bmatrix}$$

# Creating a Portfolio

➤ A portfolio is investing  $w = \begin{bmatrix} w_1 \\ \dots \\ w_N \end{bmatrix}$  in  $N$  assets.

➤ The return of the portfolio is:  $R_p = w_1 R_1 + w_2 R_2 + \dots + w_N R_N$

➤ The expected return of the portfolio is:

$$\begin{aligned} E(R_p) &= w_1 E(R_1) + w_2 E(R_2) + \dots + w_N E(R_N) = \\ &= w_1 \mu_1 + w_2 \mu_2 + \dots + w_N \mu_N = \sum_{i=1}^N w_i \mu_i = w' \mu \end{aligned}$$

# Creating a Portfolio

The variance of the portfolio is:

$$\begin{aligned}\sigma_p^2 = \text{VAR}(R_p) &= w_1^2\sigma_1^2 + w_1w_2\sigma_{12} + \cdots w_1w_N\sigma_{1N} \\ &+ w_1w_2\sigma_{12} + w_2^2\sigma_2^2 + \cdots + w_2w_N\sigma_{2N} \\ &+ \\ &\vdots \\ &+ \\ &+ w_1w_N\sigma_{1N} + w_2w_N\sigma_{2N} + \cdots + w_N^2\sigma_N^2\end{aligned}$$

# Creating a Portfolio

- Thus  $\sigma_p^2 = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_i \sigma_j \rho_{ij}$   
where  $\rho_{ij}$  is the coefficient of correlation.
- Using matrix notation:  $\sigma_p^2 = \underset{1 \times 1}{w'} \underset{1 \times N}{V} \underset{N \times N}{w}$

# The Case of Two Risky Assets

➤ To illustrate, let us consider two risky assets:

$$R_p = w_1 R_1 + w_2 R_2$$

➤ We know:  $\sigma_p^2 = VAR(R_{pt}) = w_1^2 \sigma_1^2 + 2w_1 w_2 \sigma_{12} + w_2^2 \sigma_2^2$

➤ Let us check:

$$\sigma_p^2 = (w_1, w_2) \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = w_1^2 \sigma_1^2 + 2w_1 w_2 \sigma_{12} + w_2^2 \sigma_2^2$$

➤ So it works!

# Dominance of the Covariance

- When the number of assets is large, the covariances define the portfolio's volatility.
- To illustrate, assume that all assets have the same volatility and the same pairwise correlations.
- Then an equal weight portfolio's variation is

$$\sigma_p^2 = \sigma^2 \left[ \frac{N + \rho N(N - 1)}{N^2} \right] \xrightarrow{N \rightarrow \infty} \sigma^2 \rho = \text{cov}$$

# The Efficient Frontier: Excluding Risk-free Asset

The optimization program:

$$\begin{aligned} \min \quad & w'Vw \\ \text{s.t.} \quad & w'\iota = 1 \\ & w'\mu = \mu_p \end{aligned}$$

where  $\iota$  is the Greek letter iota  $\iota_{N \times 1} = \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix}$ ,

and where  $\mu_p$  is the expected return target set by the investor.



# The Efficient Frontier: Excluding Risk-free Asset

Using the Lagrange setup:

$$L = \frac{1}{2}w'Vw + \lambda_1(1 - w'\iota) + \lambda_2(\mu_p - w'\mu)$$

$$\frac{\partial L}{\partial w} = Vw - \lambda_1\iota - \lambda_2\mu = 0$$

$$w = \lambda_1 V^{-1}\iota + \lambda_2 V^{-1}\mu$$

# The Efficient Frontier: Without Risk-free Asset

► Let

$$a = \iota'V^{-1}\mu$$

$$b = \mu'V^{-1}\mu$$

$$c = \iota'V^{-1}\iota$$

$$d = bc - a^2$$

$$g = \frac{1}{d}(bV^{-1}\iota - aV^{-1}\mu)$$

$$h = \frac{1}{d}(cV^{-1}\mu - aV^{-1}\iota)$$

► The optimal portfolio is:  $w_{N \times 1}^* = g_{N \times 1} + h_{N \times 1}\mu_p$

# Examine the Optimal Solution

- That is, once you specify the expected return target, the optimal portfolio follows immediately.
- Let us check whether the sum of weights is equal to 1.

# Examine the Optimal Solution

$$l'w = l'g + l'h\mu_p$$

$$l'g = \frac{1}{d} (bl'V^{-1}l - al'V^{-1}\mu) = \frac{1}{d} (bc - a^2) = \frac{d}{d} = 1$$

$$l'h = \frac{1}{d} (cl'V^{-1}\mu - al'V^{-1}l) = \frac{1}{d} (ac - ac) = 0$$

Indeed,  $w'l = 1$ .

# Examine the Optimal Solution

➤ Let us now check whether the expected return on the portfolio is equal to  $\mu_p$ .

➤ Recall:  $w = g + h\mu_p$   
 $w'\mu = g'\mu + \mu_p(h'\mu)$

➤ So we need to show  $g'\mu = 0$   
 $h'\mu = 1$

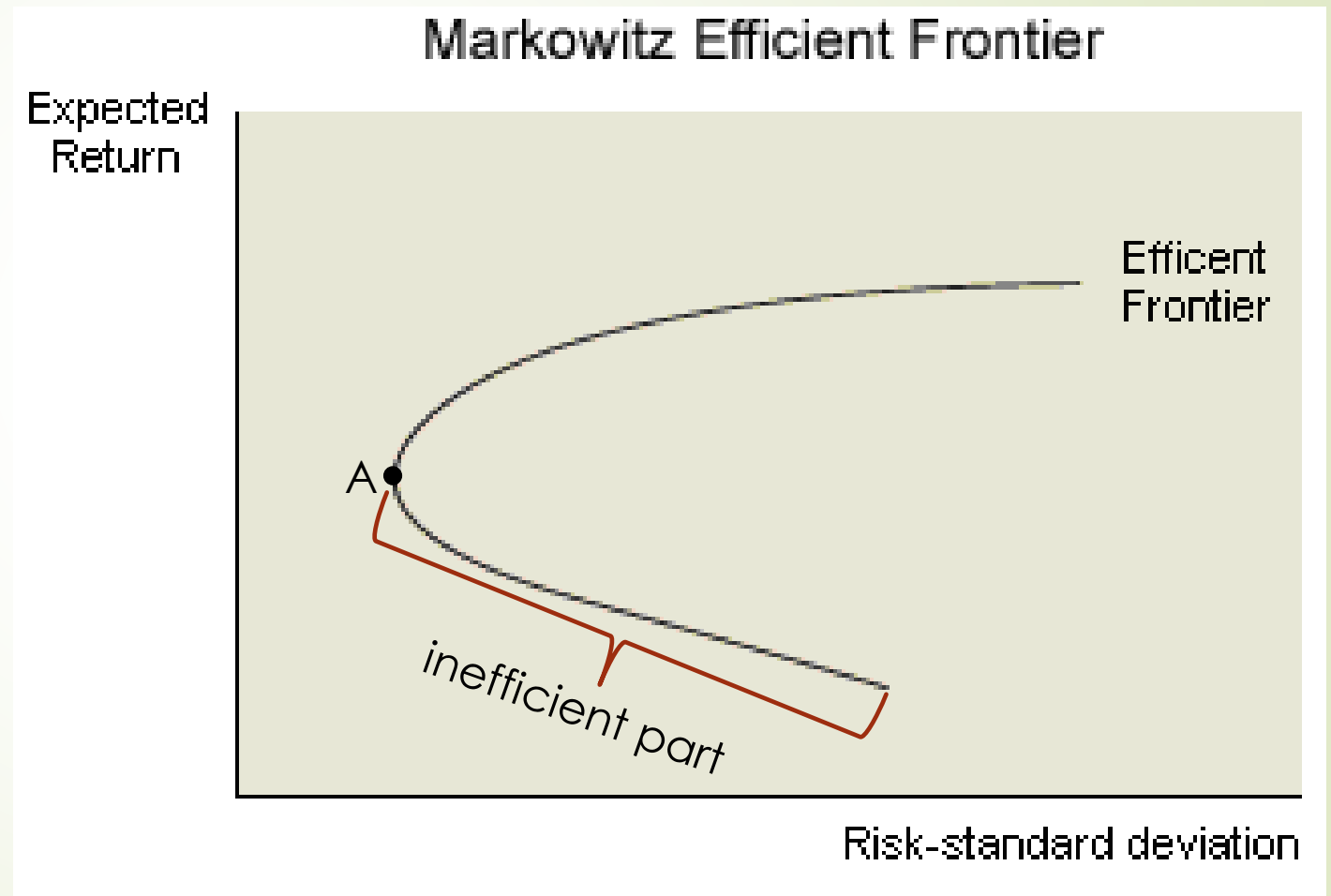
# Examine the Optimal Solution

$$g' \mu = \frac{1}{d} (b' V^{-1} \mu - a \mu' V^{-1} \mu) = \frac{1}{d} (ab - ab) = 0$$

Try it yourself: prove that  $h' \mu = 1$ .

# Efficient Frontier

The optimization program also delivers the shape of the frontier.





# Efficient Frontier

- Point A stands for the Global Minimum Variance Portfolio (GMVP).
- The efficient frontier reflects the investment opportunities; this is the supply side in partial equilibrium.
- Points below A are inefficient since they are being dominated by other portfolios that deliver better risk-return tradeoff.

# The Notion of Dominance

If

$$\sigma_B \leq \sigma_A$$
$$\mu_B \geq \mu_A$$

and there is at least one strong inequality, then portfolio B dominates portfolio A.

# The Efficient Frontier with Risk-free Asset

In practice, there is not really a risk-free asset. Why?

- Credit risk (see Greece, Spain, Iceland).
- Inflation risk.
- Interest rate risk and re-investment risk when the investment horizon is longer than the maturity of the supposedly risk-free instrument or when coupons are paid.

# Setting the Optimization

## in the Presence of Risk-free Asset

- The optimal solution is given by:

$$\min w'Vw$$

$$\text{s.t } w'\mu + (1 - w'\iota)R_f = \mu_p \quad \text{or,} \quad R_f + w'\mu^e = \mu_p$$

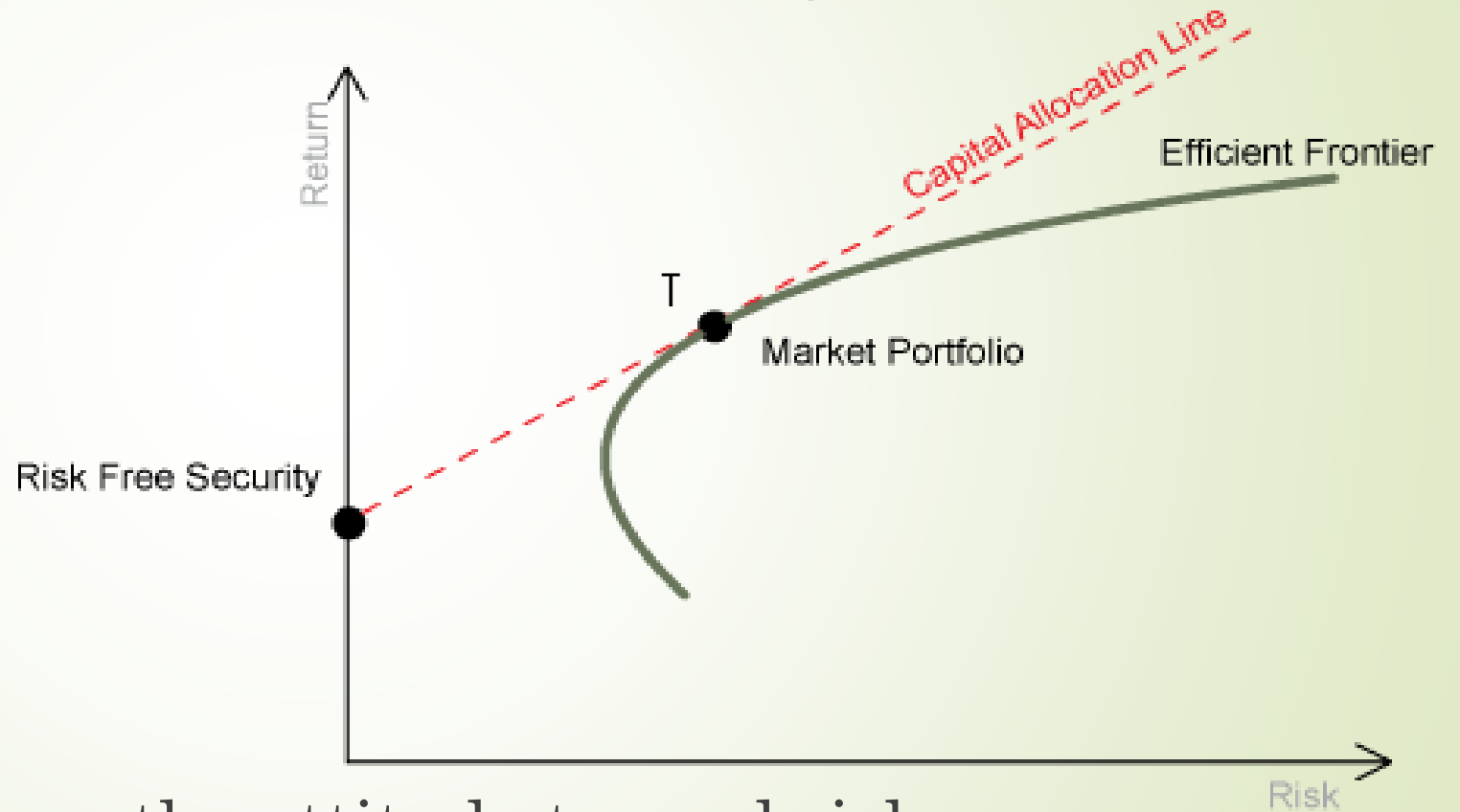
and the tangency portfolio takes the form:

$$\bar{w}^* = \frac{V^{-1}(\mu - R_f \cdot \iota)}{\iota'V^{-1}(\mu - R_f \cdot \iota)} = \frac{V^{-1}\mu^e}{\iota'V^{-1}\mu^e}$$

- The tangency portfolio is investing all the funds in risky assets.

# The Investment Opportunities

- However, the investor could select any point in the line emerging from the risk-free rate and touching the efficient frontier in point  $T$ .



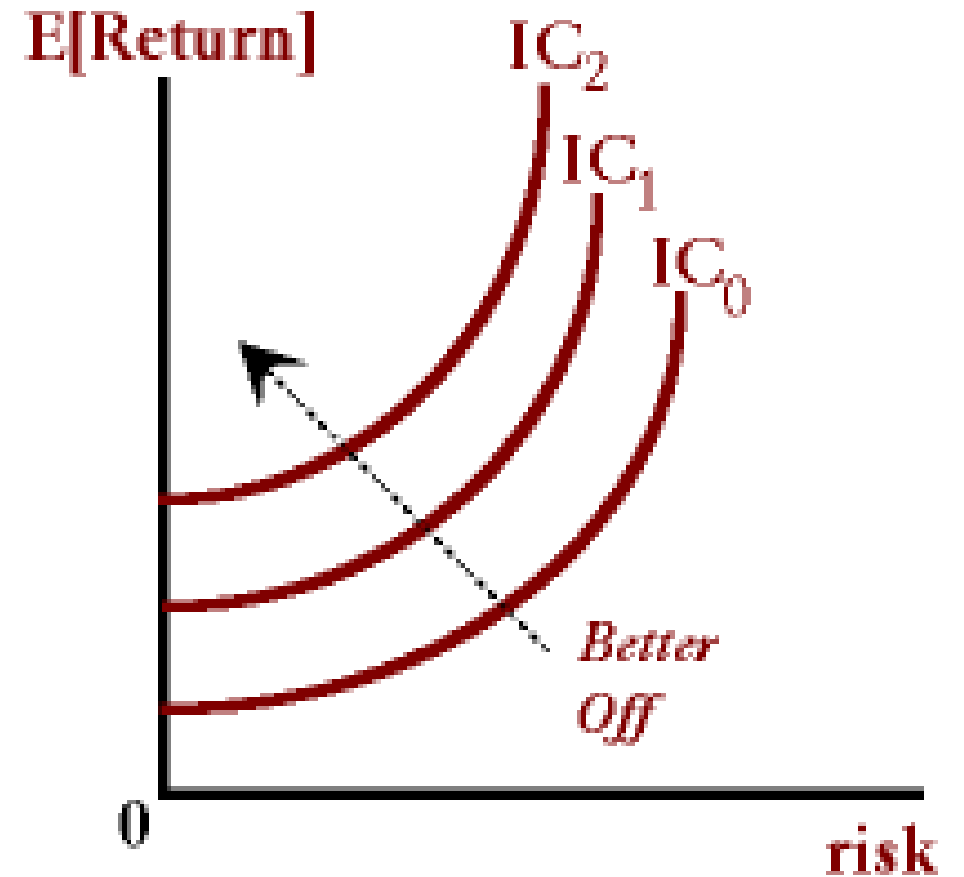
- The location depends on the attitude toward risk.

# Fund Separation

- Interestingly, all investors in the economy will mix the tangency portfolio and the risk-free asset.
- The mix depends on preferences.
- But the proportion of risky assets will be equal across the board.
- One way to test the CAPM is indeed to examine whether all investors hold the same proportions of risky assets.
- Obviously they don't!

# Equilibrium

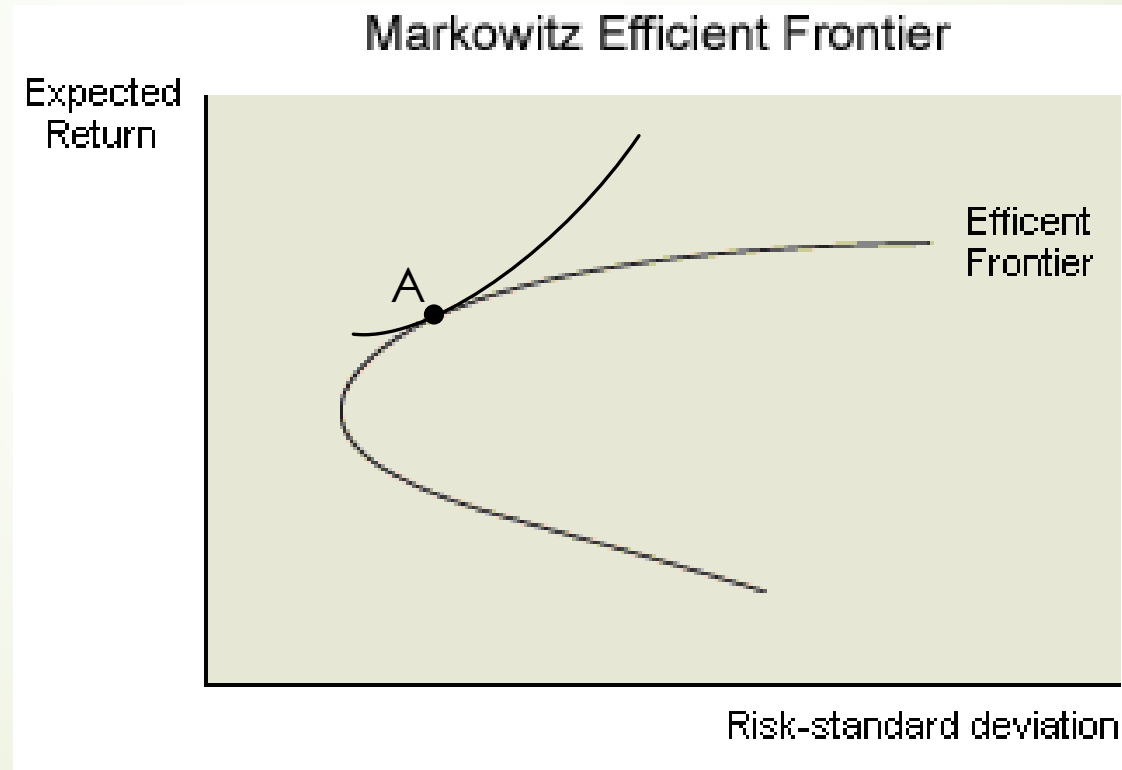
- The efficient frontier reflects the supply side.
- What about the demand?
- The demand side can be represented by a set of indifference curves.





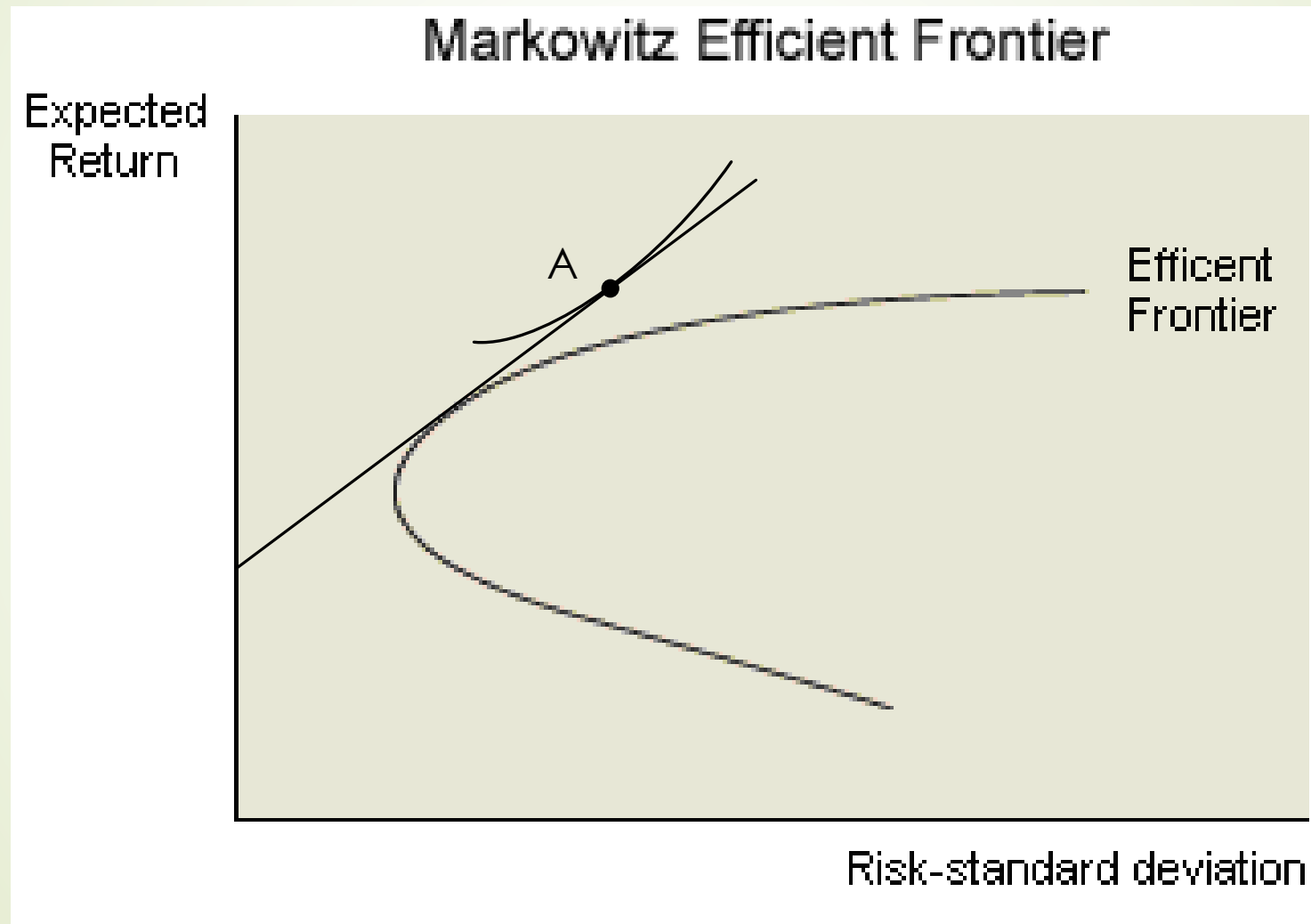
# Equilibrium

- What is the slope of indifference curve positive?
- The equilibrium obtains when the indifference curve tangents the efficient frontier
- No risk-free asset:



# Equilibrium

With risk-free asset:



# Maximize Utility / Certainty Equivalent Return

In the presence of a risk-free security, the tangency point can be found by maximizing a utility function of the form

$$U = \mu_p - \frac{1}{2} \gamma \sigma_p^2$$

where  $\gamma$  is the relative risk aversion.

# Maximize Utility / Certainty Equivalent Return

- Notice that utility is equal to expected return minus a penalty factor.
- The penalty factor positively depends on the risk aversion (demand) and the variance (supply).

# Utility Maximization

$$U(w) = R_f + w' \mu^e - \frac{1}{2} \gamma \cdot w' V w$$

$$\frac{\partial U}{\partial w} = \mu^e - \gamma V w = 0$$

$$w^* = \frac{1}{\gamma} V^{-1} \mu^e$$

The utility maximization yields the same tangency portfolio

$$\bar{w}^* = \frac{V^{-1} \mu_e}{\iota' V^{-1} \mu_e}$$

where  $w^*$  reflects the fraction of weights invested in risky assets. The rest is invested in the risk-free asset.

# Mixing the Risky and Risk-free Assets

$$w^* = \frac{1}{\gamma} V^{-1} \mu^e$$

- So if  $\gamma = \iota' V^{-1} \mu^e$  then all funds (100%) are invested in risky assets.
- If  $\gamma > \iota' V^{-1} \mu^e$  then some fraction is invested in risk-free asset.
- If  $\gamma < \iota' V^{-1} \mu^e$  then the investor borrows money to leverage his/her equity position.

# The Exponential Utility Function

- The exponential utility function is of the form

$$U(W) = -\exp(-\lambda W) \text{ where } \lambda > 0$$

- Notice that

$$U'(W) = \lambda \exp(-\lambda W) > 0$$
$$U''(W) = -\lambda^2 \exp(-\lambda W) < 0$$

- That is, the marginal utility is positive but it diminishes with an increasing wealth.



# The Exponential Utility Function

$$ARA = -\frac{U''}{U'} = \lambda$$

- Indeed, the exponential preferences belong to the class of constant absolute risk aversion (CARA).
- For comparison, power preferences belong to the class of constant relative risk aversion (CRRA).

# Exponential: The Optimization Mechanism

➤ The investor maximizes the expected value of the exponential utility where the decision variable is the set of weights  $w$  and subject to the wealth evolution.

➤ That is

$$\max_w E[U(W_{t+1})|W_t]$$

$$s. t. \quad W_{t+1} = W_t(1 + R_f + w'R_{t+1}^e)$$

# Exponential: The Optimization Mechanism

► Let us assume that  $R_{t+1}^e \sim N(\mu^e, V)$

► Then  $W_{t+1} \sim N \left[ \underbrace{W_t(1 + R_f + w'\mu^e)}_{\text{mean}}, \underbrace{W_t^2 w' V w}_{\text{variance}} \right]$

# Exponential: The Optimization Mechanism

It is known that for  $x \sim N(\mu_x, \sigma_x^2)$

$$E[\exp(ax)] = \exp\left(a\mu_x + \frac{1}{2}a^2\sigma_x^2\right)$$

# Exponential: The Optimization Mechanism

Thus,

$$\begin{aligned} -E[\exp(-\lambda W_{t+1})] &= -\exp\left(-\lambda E(W_{t+1}) + \frac{1}{2}\lambda^2 \cdot \text{VAR}(W_{t+1})\right) \\ &= -\exp\left(-\lambda W_t(1 + R_f + w'\mu^e) + \frac{1}{2}\lambda^2 W_t^2 w'Vw\right) \\ &= -\exp\left(-\lambda W_t(1 + R_f)\right) \exp\left(-\lambda W_t\left(w'\mu^e - \frac{1}{2}\lambda W_t w'Vw\right)\right) \end{aligned}$$

# Exponential: The Optimization Mechanism

Notice that  $\gamma = \lambda W_t$

where  $\gamma$  is the relative risk aversion coefficient.

# Exponential: The Optimization Mechanism

- So the investor ultimately maximizes

$$\max_w \left[ w' \mu^e - \frac{1}{2} \gamma w' V w \right]$$

- The optimal solution is  $w^* = \frac{1}{\gamma} V^{-1} \mu^e$ .

- The tangency portfolio is the same as before

$$\bar{w}^* = \frac{V^{-1} \mu^e}{l' V^{-1} \mu^e}$$



# Exponential: The Optimization Mechanism

## Conclusion:

The joint assumption of exponential utility and normally distributed stock return leads to the well-known mean variance solution.

# Quadratic Preferences

► The quadratic utility function is of the form

$$U(W) = a + W - \frac{b}{2}W^2 \text{ where } b > 0$$

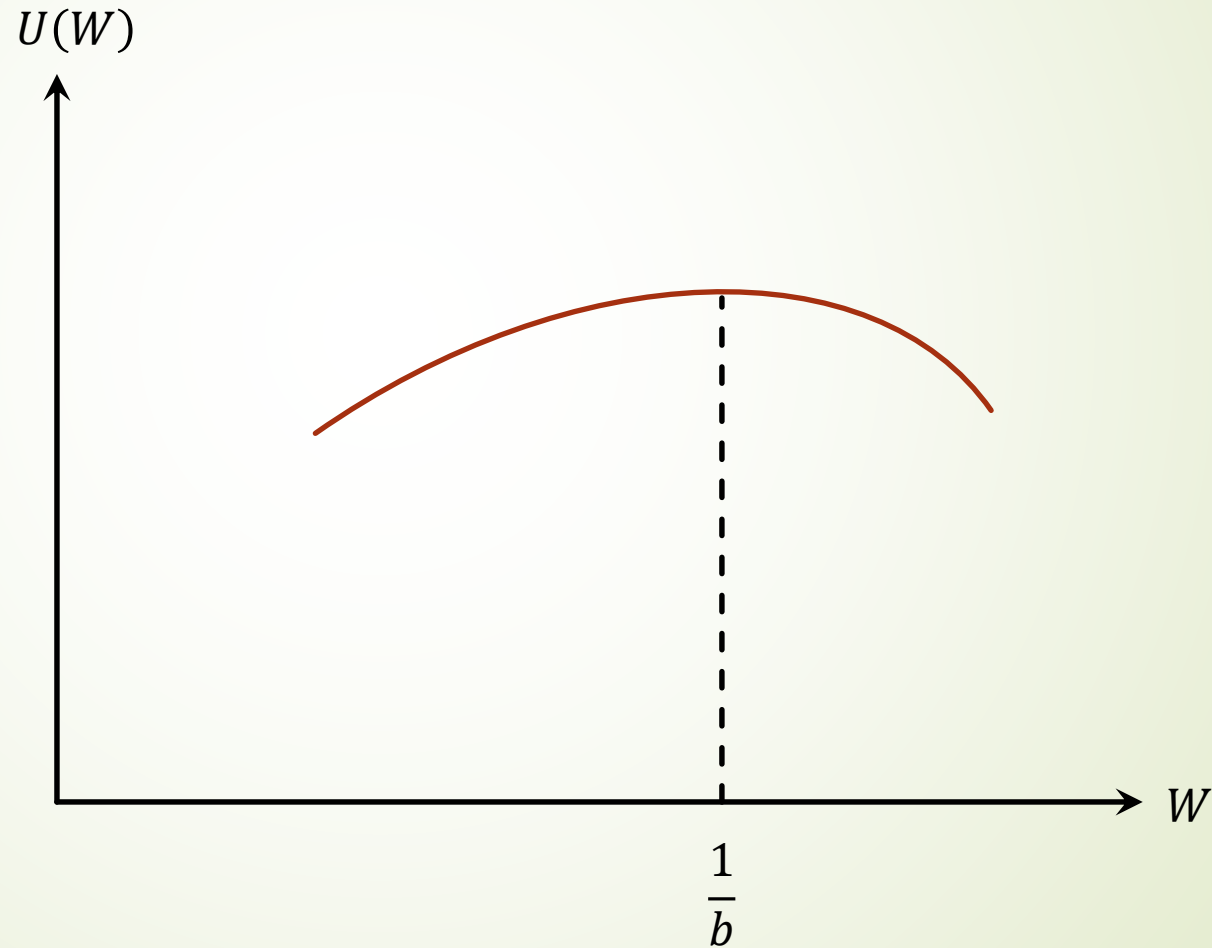
$$\frac{\partial U}{\partial W} = 1 - bW$$

$$\frac{\partial^2 U}{\partial W^2} = -b < 0$$

► Notice that the first derivative is positive for  $b < \frac{1}{W}$

# Quadratic Preferences

The utility function looks like



# Quadratic Preferences

➤ It has a diminishing part – which makes no sense – because we always prefer higher than lower wealth

➤ Utility is thus restricted to the positive slope part

➤ Notice that  $\gamma = RRA = -\frac{U''}{U'} W = \frac{bW}{1-bW}$

➤ The optimization formulation is given by

$$\max_w E[U(W_{t+1})|W_t]$$

$$s. t. W_{t+1} = W_t(1 + R_{ft} + w'R_{t+1}^e)$$

# Quadratic Preferences

- Avramov and Chordia (2006 JFE) show that the optimization could be formulated as

$$\max_w w' \mu^e - \frac{1}{2 \left( \frac{1}{\gamma} - R_{ft} \right)} w' \left( V + \mu^e \mu^{e'} \right)^{-1} w$$

- The solution takes the form

$$w^* = \left( \frac{1}{\gamma} - R_{ft} \right) \frac{V^{-1} \mu^e}{1 + \mu^{e'} V^{-1} \mu^e}$$

# Quadratic Preferences

- The tangency portfolio is

$$\bar{w}^* = \frac{V^{-1}\mu^e}{\iota'V^{-1}\mu^e}$$

- The only difference from previously presented competing specifications is the composition of risky and risk-free assets.
- But in all solutions the proportions of risky assets are identical.

# The Sharpe Ratio of the Tangency Portfolio

► Notice that  $\mu^{e'} V^{-1} \mu^e$  is actually the squared Sharpe Ratio of the tangency portfolio.

► Let us prove it

$$\bar{w}^* = \frac{V^{-1} \mu^e}{\iota' V^{-1} \mu^e}$$

$$\mu_{TP} - R_f = \bar{w}^{*'} \mu^e + (1 - \bar{w}^{*'} \iota) R_f = \bar{w}^* \mu^e = \frac{\mu^{e'} V^{-1} \mu^e}{\iota' V^{-1} \mu^e}$$

$$\sigma_{TP}^2 = \bar{w}^{*'} V \bar{w}^* = \frac{\mu^{e'} V^{-1} \mu^e}{(\iota' V^{-1} \mu^e)^2}$$



# The Sharpe Ratio of the Tangency Portfolio

➤ Thus,  $SR_{TP}^2 = \frac{(\mu_{TP} - R_f)^2}{\sigma_{TP}^2} = \mu^e V^{-1} \mu^e$

➤  $TP$  is a subscript for the tangency portfolio.

# Session #5: Testing Asset Pricing Models: Time Series Perspective

# Why Caring about Asset Pricing Models?

- An essential question that arises is why would both academics and practitioners invest huge resources in developing and testing asset pricing models.
- It turns out that pricing models have crucial roles in various applications in financial economics – both asset pricing as well as corporate finance.
- In the following, I list five major applications.

# 1 – Common Risk Factors

- Pricing models characterize the risk profile of a firm.
- In particular, systematic risk is no longer stock return volatility – rather it is the loadings on risk factors.
- For instance, in the single factor CAPM the market beta – or the co-variation with the market – characterizes the systematic risk of a firm.

# 1 – Common Risk Factors

- Likewise, in the single factor (C)CAPM the consumption growth beta – or the co-variation with consumption growth – characterizes the systematic risk of a firm.
- In the multi-factor Fama-French (FF) model there are three sources of risk – the market beta, the SMB beta, and the HML beta.

# 1 – Common Risk Factors

- Under FF, other things being equal (*ceteris paribus*), a firm is riskier if its loading on SMB beta is higher.
- Under FF, other things being equal (*ceteris paribus*), a firm is riskier if its loading on HML beta is higher.

## 2 – Moments for Asset Allocation

- Pricing models deliver moments for asset allocation.
- For instance, the tangency portfolio takes on the form

$$w_{TP} = \frac{V^{-1}\mu^e}{l'V^{-1}\mu^e}$$



## 2 – Asset Allocation

Under the CAPM, the vector of expected returns and the covariance matrix are given by:

$$\begin{aligned}\mu^e &= \beta \mu_m^e \\ V &= \beta \beta' \sigma_m^2 + \Sigma\end{aligned}$$

where  $\Sigma$  is the covariance matrix of the residuals in the time-series asset pricing regression.

We denoted by  $\Psi$  the residual covariance matrix in the case wherein the off diagonal elements are zeroed out.

## 2 – Asset Allocation

The corresponding quantities under the FF model are

$$\mu^e = \beta_{MKT}\mu_m^e + \beta_{SML}\mu_{SML} + \beta_{HML}\mu_{HML}$$

$$V = \beta\Sigma_F\beta' + \Sigma$$

where  $\Sigma_F$  is the covariance matrix of the factors.

# 3 – Discount Factors

- ▶ Expected return is the discount factor, commonly denoted by  $k$ , in present value formulas in general and firm evaluation in particular:

$$PV = \sum_{t=1}^T \frac{CF_t}{(1+k)^t}$$

- ▶ In practical applications, expected returns are typically assumed to be constant over time, an unrealistic assumption.

## 3 – Discount Factors

- Indeed, thus far we have examined models with constant beta and constant risk premiums

$$\mu^e = \beta' \lambda$$

where  $\lambda$  is a  $K$ -vector of risk premiums.

- When factors are return spreads the risk premium is the mean of the factor.
- Later we will consider models with time varying factor loadings.

## 4 – Benchmarks

- Factors in asset pricing models serve as benchmarks for evaluating performance of active investments.
- In particular, performance is the intercept (alpha) in the time series regression of excess fund returns on a set of benchmarks (typically four benchmarks in mutual funds and more so in hedge funds):

$$r_t^e = \alpha + \beta_{MKT} \times r_{MKT,t}^e + \beta_{SMB} \times SMB_t + \beta_{HML} \times HML_t + \beta_{WML} \times WML_t + \varepsilon_t$$

# 5 – Corporate Finance

- There is a plethora of studies in corporate finance that use asset pricing models to risk adjust asset returns.
- Here are several examples:
  - Examining the long run performance of IPO firm.
  - Examining the long run performance of SEO firms.
  - Analyzing abnormal performance of stocks going through dividend initiation, dividend omission, splits, and reverse splits.

# 5 – Corporate Finance

- ▶ Analyzing mergers and acquisitions
- ▶ Analyzing the impact of change in board of directors.
- ▶ Studying the impact of corporate governance on the cross section of average returns.
- ▶ Studying the long run impact of stock/bond repurchase.



# Time Series Tests

- ▶ Time series tests are designated to examine the validity of asset pricing models in which factors are return spreads.
- ▶ Example: the market factor is the return difference between the market portfolio and the risk-free asset.
- ▶ Consumption growth is not a return spread.
- ▶ Thus, the consumption CAPM cannot be tested using time series regressions, unless you form a factor mimicking portfolio (FMP) for consumption growth.

# Time Series Tests

- FMP is a convex combination of returns on some basis assets having the maximal correlation with consumption growth.
- The statistical time series tests have an appealing economic interpretation. In particular:
- Testing the CAPM amounts to testing whether the market portfolio is the tangency portfolio.
- Testing multi-factor models amounts to testing whether some particular combination of the factors is the tangency portfolio.

# Testing the CAPM

- ▶ Run the time series regression:

$$\begin{aligned}r_{1t}^e &= \alpha_1 + \beta_1 r_{mt}^e + \varepsilon_{1t} \\ &\vdots \\ r_{Nt}^e &= \alpha_N + \beta_N r_{mt}^e + \varepsilon_{Nt}\end{aligned}$$

- ▶ The null hypothesis is:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_N = 0$$

# Testing the CAPM

- In the following, I will introduce four times series test statistics:
  - WALD.
  - Likelihood Ratio.
  - GRS (Gibbons, Ross, and Shanken (1989)).
  - GMM.

# The Distribution of $\alpha$ .

➤ Recall,  $\alpha$  is asset mispricing.

➤ The time series regressions can be rewritten using a vector form as:

$$\underset{N \times 1}{r_t^e} = \underset{N \times 1}{\alpha} + \underset{N \times 1}{\beta} \cdot \underset{1 \times 1}{r_{mt}^e} + \underset{N \times 1}{\varepsilon_t}$$

➤ Let us assume that

$$\underset{N \times 1}{\varepsilon_t} \stackrel{iid}{\sim} N \left( 0, \underset{N \times N}{\Sigma} \right) \text{ for } t = 1, 2, 3, \dots, T$$

➤ Let  $\theta = (\alpha', \beta', \text{vech}(\Sigma)')$  be the set of all parameters.

# The Distribution of $\alpha$ .

► Under normality, the likelihood function for  $\varepsilon_t$  is

$$L(\varepsilon_t|\theta) = c \Sigma^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (r_t^e - \alpha - \beta r_{mt}^e)' \Sigma^{-1} (r_t^e - \alpha - \beta r_{mt}^e) \right]$$

where  $c$  is the constant of integration (recall the integral of a probability distribution function is unity).

# The Distribution of $\alpha$ .

➤ Moreover, the IID assumption suggests that

$$L(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N | \theta) = c^T \Sigma^{-\frac{T}{2}}$$

$$\times \exp \left[ -\frac{1}{2} \sum_{t=1}^T (r_t^e - \alpha - \beta r_{mt}^e)' \Sigma^{-1} (r_t^e - \alpha - \beta r_{mt}^e) \right]$$

➤ Taking the natural log from both sides yields

$$\ln(L) \propto -\frac{T}{2} \ln(\Sigma) - \frac{1}{2} \sum_{t=1}^T (r_t^e - \alpha - \beta r_{mt}^e)' \Sigma^{-1} (r_t^e - \alpha - \beta r_{mt}^e)$$



# The Distribution of $\alpha$ .

Asymptotically, we have  $\theta - \hat{\theta} \sim N(0, \Sigma(\theta))$

where

$$\Sigma(\theta) = \left[ -E \left[ \frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \right] \right]^{-1}$$

# The Distribution of $\alpha$ .

Let us estimate the parameters

$$\frac{\partial \ln(L)}{\partial \alpha} = \Sigma^{-1} \left[ \sum_{t=1}^T (r_t^e - \alpha - \beta r_{mt}^e) \right]$$

$$\frac{\partial \ln(L)}{\partial \beta} = \Sigma^{-1} \left[ \sum_{t=1}^T (r_t^e - \alpha - \beta r_{mt}^e) \times r_{mt}^e \right]$$

$$\frac{\partial \ln(L)}{\partial \Sigma} = -\frac{T}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \left[ \sum_{t=1}^T \varepsilon_t \varepsilon_t' \right] \Sigma^{-1}$$

# The Distribution of $\alpha$ .

Solving for the first order conditions yields

$$\hat{\alpha} = \hat{\mu}^e - \hat{\beta} \cdot \hat{\mu}_m^e$$

$$\hat{\beta} = \frac{\sum_{t=1}^T (r_t^e - \hat{\mu}^e) (r_{mt}^e - \hat{\mu}_m^e)}{\sum_{t=1}^T (r_{mt}^e - \hat{\mu}_m^e)^2}$$

# The Distribution of $\alpha$ .

Moreover,

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

$$\hat{\mu}^e = \frac{1}{T} \sum_{t=1}^T r_t^e$$

$$\mu_m^e = \frac{1}{T} \sum_{t=1}^T r_{mt}^e$$

# The Distribution of $\alpha$ .

- Recall our objective is to find the standard errors of  $\hat{\alpha}$ .
- Standard errors could be found using the information matrix.

# The Distribution of $\alpha$ .

► The information matrix is constructed as follows

$$I(\theta) = - \left\{ E \left[ \begin{array}{ccc} \frac{\partial^2 \ln(L)}{\partial \alpha \partial \alpha'} & \frac{\partial^2 \ln(L)}{\partial \alpha \partial \beta'} & \frac{\partial^2 \ln(L)}{\partial \alpha \partial \Sigma'} \\ \frac{\partial^2 \ln(L)}{\partial \beta \partial \alpha'} & \frac{\partial^2 \ln(L)}{\partial \beta \partial \beta'} & \frac{\partial^2 \ln(L)}{\partial \beta \partial \Sigma'} \\ \frac{\partial^2 \ln(L)}{\partial \Sigma \partial \alpha'} & \frac{\partial^2 \ln(L)}{\partial \Sigma \partial \beta'} & \frac{\partial^2 \ln(L)}{\partial \Sigma \partial \Sigma'} \end{array} \right] \right\}$$

# The Distribution of the Parameters

- Exercise: establish the information matrix.
- Notice that  $\hat{\alpha}$  and  $\hat{\beta}$  are independent of  $\hat{\Sigma}$  - thus, you can ignore the second derivatives with respect to  $\Sigma$  in the information matrix if your objective is to find the distribution of  $\hat{\alpha}$  or  $\hat{\beta}$  or both.
- If you aim to derive the distribution of  $\hat{\Sigma}$  then focus on the bottom right block of the information matrix.



# The Distribution of $\alpha$ .

➔ We get:

$$\hat{\alpha} \sim N \left( \alpha, \frac{1}{T} \left[ 1 + \left( \frac{\hat{\mu}_m^e}{\hat{\sigma}_m} \right)^2 \right] \Sigma \right)$$

➔ Moreover,

$$\hat{\beta} \sim N \left( \beta, \frac{1}{T} \cdot \frac{1}{\hat{\sigma}_m^2} \Sigma \right)$$

$$T\hat{\Sigma} \sim W(T - 2, \Sigma)$$

# The Distribution of $\alpha$ .

Notice that  $W(x, y)$  stands for the Wishart distribution with  $x = T - 2$  degrees of freedom and a parameter matrix  $y = \Sigma$ .

# The Wald Test

➤ Recall, if

$$X \sim N(\mu, \Sigma) \quad \text{then} \quad (X - \mu)' \hat{\Sigma}^{-1} (X - \mu) \sim \chi^2(N)$$

➤ Here we test

$$H_0: \hat{\alpha} = 0 \quad \text{where} \quad \hat{\alpha} \stackrel{H_0}{\sim} N(0, \Sigma_\alpha)$$
$$H_1: \hat{\alpha} \neq 0$$

➤ The Wald statistic is  $\hat{\alpha}' \hat{\Sigma}_\alpha^{-1} \hat{\alpha} \sim \chi^2(N)$

# The Wald Test

which becomes:

$$J_1 = T \left[ 1 + \left( \frac{\hat{\mu}_m^e}{\hat{\sigma}_m} \right)^2 \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} = T \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + S\hat{R}_m^2}$$

where  $S\hat{R}_m$  is the Sharpe ratio of the market factor.

# Algorithm for Implementation

- The algorithm for implementing the statistic is as follows:
- Run separate regressions for the test assets on the common factor:

$$\begin{array}{l} r_1^e \\ T \times 1 \end{array} = \begin{array}{l} X \\ T \times 2 \end{array} \begin{array}{l} \theta_1 \\ 2 \times 1 \end{array} + \begin{array}{l} \varepsilon_1 \\ T \times 1 \end{array}$$

⋮

$$\begin{array}{l} r_N^e \\ T \times 1 \end{array} = \begin{array}{l} X \\ T \times 2 \end{array} \begin{array}{l} \theta_N \\ 2 \times 1 \end{array} + \begin{array}{l} \varepsilon_N \\ T \times 1 \end{array}$$

where

$$X_{T \times 2} = \begin{bmatrix} 1, r_{m1}^e \\ \vdots \\ 1, r_{mT}^e \end{bmatrix}$$
$$\theta_i = [\alpha_i, \beta_i]'$$

# Algorithm for Implementation

- Retain the estimated regression intercepts

$$\hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_N]' \quad \text{and} \quad \hat{\varepsilon}_{T \times N} = [\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N]$$

- Compute the residual covariance matrix

$$\hat{\Sigma} = \frac{1}{T} \hat{\varepsilon}' \hat{\varepsilon}$$

# Algorithm for Implementation

- Compute the sample mean and the sample variance of the factor.
- Compute  $J_1$ .



# The Likelihood Ratio Test

► We run the unrestricted and restricted specifications:

$$\text{un: } r_t^e = \alpha + \beta r_{mt}^e + \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma)$$

$$\text{res: } r_t^e = \beta^* r_{mt}^e + \varepsilon_t^* \quad \varepsilon_t^* \sim N(0, \Sigma^*)$$

► Using MLE, we get:

# The Likelihood Ratio Test

$$\hat{\beta}^* = \frac{\sum_{t=1}^T r_t^e r_{mt}^e}{\sum_{t=1}^T (r_{mt}^e)^2}$$

$$\hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^* \hat{\varepsilon}_t^{*'}$$

$$\hat{\beta}^* \sim N \left( \beta, \frac{1}{T} \left[ \frac{1}{(\hat{\mu}_m^e)^2 + \hat{\sigma}_m^2} \right] \Sigma \right)$$

$$T\hat{\Sigma}^* \sim W(T - 1, \Sigma)$$

# The LR Test

$$LR = \ln(L^*) - \ln(L) = -\frac{T}{2} [\ln |\hat{\Sigma}^*| - \ln |\hat{\Sigma}|]$$

$$J_2 = -2LR = T [\ln |\hat{\Sigma}^*| - \ln |\hat{\Sigma}|] \sim \chi^2(N)$$

► Using some algebra, one can show that

$$J_1 = T \left( \exp \left( \frac{J_2}{T} \right) - 1 \right)$$

► Thus,

$$J_2 = T \cdot \ln \left( \frac{J_1}{T} + 1 \right)$$

# GRS (1989)

Theorem: let  $X_{N \times 1} \sim N(0, \Sigma)$

let  $A_{N \times N} \sim W(\tau, \Sigma)$  where  $\tau \geq N$

and let  $A$  and  $X$  be independent then

$$\frac{\tau - N + 1}{N} X' A^{-1} X \sim F_{N, \tau - N + 1}$$

# GRS (1989)

In our context:

$$X = \sqrt{T} \left[ 1 + \left( \frac{\hat{\mu}_m^e}{\hat{\sigma}_m} \right)^2 \right]^{-\frac{1}{2}} \hat{\alpha} \stackrel{H_0}{\sim} N(0, \Sigma)$$
$$A = T\hat{\Sigma} \sim W(\tau, \Sigma)$$

where

$$\tau = T - 2$$

Then:

$$J_3 = \left( \frac{T - N - 1}{N} \right) \left[ 1 + \left( \frac{\hat{\mu}_m^e}{\hat{\sigma}_m} \right)^2 \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F(N, T - N - 1)$$

This is a finite-sample test.

# GMM

I will directly give the statistic without derivation:

$$J_4 = T \hat{\alpha}' \left( R (D_T' S_T^{-1} D_T)^{-1} R' \right)^{-1} \cdot \hat{\alpha} \stackrel{H_0}{\sim} \chi^2 (N)$$

where

$$R_{N \times 2N} = \begin{bmatrix} I_N & \mathbf{0} \\ N \times N & N \times N \end{bmatrix}$$

$$D_T = - \begin{bmatrix} 1, & \hat{\mu}_m^e \\ \hat{\mu}_m^e, & (\hat{\mu}_m^e)^2 + \hat{\sigma}_m^2 \end{bmatrix} \otimes I_N$$

# GMM

- Assume no serial correlation but Heteroskedasticity:

$$S_T = \frac{1}{T} \sum_{t=1}^T (x_t x_t' \otimes \hat{\varepsilon}_t \hat{\varepsilon}_t')$$

where

$$x_t = [1, r_{mt}^e]'$$

- Under homoscedasticity and serially uncorrelated moment conditions:  $J_4 = J_1$ .
- That is, the GMM statistic boils down to the WALD.



# The Multi-Factor Version of Asset Pricing Tests

$$r_t^e = \alpha + \beta \cdot F_t + \varepsilon_t$$

$N \times 1$        $N \times 1$        $N \times K$        $K \times 1$        $N \times 1$

$$J_1 = T(1 + \hat{\mu}'_F \hat{\Sigma}_F^{-1} \hat{\mu}_F)^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim \chi(N)$$

$J_2$  follows as described earlier.

$$J_3 = \frac{T - N - K}{N} (1 + \hat{\mu}'_F \hat{\Sigma}_F^{-1} \hat{\mu}_F)^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F_{(N, T-N-K)}$$

where  $\hat{\mu}_F$  is the mean vector of the factor based return spreads.

# The Multi-Factor Version of Asset Pricing Tests

➔  $\hat{\Sigma}_F$  is the variance covariance matrix of the factors.

➔ For instance, considering the Fama-French model:

$$\hat{\mu}_F = \begin{bmatrix} \hat{\mu}_m^e \\ \hat{\mu}_{SMB} \\ \hat{\mu}_{HML} \end{bmatrix} \quad \hat{\Sigma}_F = \begin{bmatrix} \hat{\sigma}_m^2 & \hat{\sigma}_{m,SMB} & \hat{\sigma}_{m,HML} \\ \hat{\sigma}_{SMB,m} & \hat{\sigma}_{SMB}^2 & \hat{\sigma}_{SMB,HML} \\ \hat{\sigma}_{HML,m} & \hat{\sigma}_{HML,SMB} & \hat{\sigma}_{HML}^2 \end{bmatrix}$$

# The Current State of Asset Pricing Models

- The CAPM has been rejected in asset pricing tests.
- The Fama-French model is not a big success.
- Conditional versions of the CAPM and CCAPM display some improvement.
- Should decision-makers abandon a rejected CAPM?

# Should a Rejected CAPM be Abandoned?

- Often a misspecified model can improve estimation.
- In particular, assume that expected stock return is given by  $\mu_i = \alpha_i + R_f + \beta_i(\mu_m - R_f)$  where  $\alpha_i \neq 0$
- You estimate  $\mu_i$  using the sample mean and the misspecified CAPM:

$$\hat{\mu}_i^{(1)} = \frac{1}{T} \sum_{t=1}^T R_{it}$$

$$\hat{\mu}_i^{(2)} = R_f + \hat{\beta}_i(\hat{\mu}_m - R_f)$$

# Mean Squared Error (MSE)

The quality of estimates is evaluated based on the Mean Squared Error (MSE)

$$MSE^{(1)} = E \left( \hat{\mu}_i^{(1)} - \mu_i \right)^2$$

$$MSE^{(2)} = E \left( \hat{\mu}_i^{(2)} - \mu_i \right)^2$$

# MSE, Bias, and Noise of Estimates

➤ Notice that  $MSE = bias^2 + Var(estimate)$

➤ Of course, the sample mean is unbiased thus

$$MSE^{(1)} = Var\left(\hat{\mu}_i^{(1)}\right)$$

➤ However, the CAPM is rejected, thus

$$MSE^{(2)} = \alpha_i^2 + Var\left(\hat{\mu}_i^{(2)}\right)$$

# The Bias-Variance Tradeoff

It might be the case that  $Var(\hat{\mu}^{(2)})$  is significantly lower than  $Var(\hat{\mu}^{(1)})$  - thus even when the CAPM is rejected, still zeroing out  $\alpha_i$  could produce a smaller mean square error.



# When is the Rejected CAPM Superior?

$$\text{Var}(\hat{\mu}_i^{(1)}) = \frac{1}{T} \sigma^2(R_i) = \frac{1}{T} [\beta_i^2 \sigma^2(R_m) + \sigma^2(\varepsilon_i)]$$

$$\text{Var}(\hat{\mu}_i^{(2)}) = \text{Var}(\hat{\beta}_i \hat{\mu}_m^e)$$

# When is the Rejected CAPM Superior?

Using variance decomposition

$$\begin{aligned} \text{Var}(\hat{\mu}_i^{(2)}) &= \text{Var}(\hat{\beta}_i \hat{\mu}_m^e) = E[\text{Var}(\hat{\beta}_i \hat{\mu}_m^e | \hat{\mu}_m^e)] + \text{Var}(E(\hat{\beta}_i \hat{\mu}_m^e | \hat{\mu}_m^e)) \\ &= E\left[\left(\hat{\mu}_m^e\right)^2 \frac{\sigma^2(\varepsilon_i)}{\hat{\sigma}_m^2} \cdot \frac{1}{T}\right] + \text{Var}(\beta_i \hat{\mu}_m^e) = \frac{1}{T} \sigma^2(\varepsilon_i) E\left[\left(\frac{\hat{\mu}_m^e}{\hat{\sigma}_m^2}\right)^2\right] + \beta_i^2 \frac{\hat{\sigma}_m^2}{T} \\ &= \frac{1}{T} [SR_m \sigma^2(\varepsilon_i) + \beta_i^2 \hat{\sigma}_m^2] \end{aligned}$$

where

$$SR_m = E\left[\left(\frac{\hat{\mu}_m^e}{\hat{\sigma}_m^2}\right)^2\right]$$

# When is the Rejected CAPM Superior?

➔ Then

$$\frac{\text{Var}(\hat{\mu}_i^{(2)})}{\text{Var}(\hat{\mu}_i^{(1)})} = \frac{SR_m \sigma^2(\varepsilon_i) + \beta_i^2 \sigma_m^2}{\sigma^2(R_i)} = SR_m(1 - R^2) + R^2$$

where  $R^2$  is the  $R$  squared in the market regression.

➔ Since  $SR_m$  is small -- the ratio of the variance estimates is smaller than 1.

# Example

➔ Let

$$\sigma^2(R_i) = 0.01$$

$$E \left( \frac{\hat{\mu}_m^e}{\hat{\sigma}_m} \right)^2 = 0.05$$

$$R^2 = 0.3$$

- ➔ For what values of  $\alpha_i \neq 0$  it is still preferred to use the CAPM?
- ➔ Find  $\alpha_i$  such that the MSE of the CAPM is smaller.

# Example

$$\frac{MSE^{(2)}}{MSE^{(1)}} = SR_m(1 - R^2) + R^2 + \frac{\alpha_i^2}{MSE^{(1)}} < 1$$

$$0.05 \times 0.7 + 0.3 + \frac{\alpha_i^2}{\frac{1}{60} \cdot 0.01} < 1$$

$$\alpha_i^2 < \frac{1}{60} \cdot 0.01 \times 0.665$$

$$|\alpha_i| < 0.01528 = 1.528\%$$

# Economic versus Statistical Factors

- Factors such as the market portfolio, SMB, HML, WML, TERM, SPREAD are pre-specified.
- Such factors are considered to be economically based.
- For instance, Fama and French argue that SMB and HML factors are proxying for underlying state variables in the economy.

# Economic versus Statistical Factors

- Statistical factors are derived using econometric procedures on the covariance matrix of stock return.
- Two prominent methods are the factor analysis and the principal component analysis (PCA).
- Such methods are used to extract common factors.
- The first factor typically has a strong (about 96%) correlation with the market portfolio.
- Later, I will explain the PCA.



# The Economics of Time Series Test Statistics

Let us summarize the first three test statistics:

$$J_1 = T \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + S\hat{R}_m^2}$$

$$J_2 = T \cdot \ln \left( \frac{J_1}{T} + 1 \right)$$

$$J_3 = \frac{T - N - 1}{N} \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + S\hat{R}_m^2}$$

# The Economics of Time Series Test Statistics

➤ The  $J_4$  statistic, the GMM based asset pricing test, is actually a Wald test, just like  $J_1$ , except that the covariance matrix of asset mispricing takes account of heteroscedasticity and often incorporates potential serial correlation.

➤ Notice that all test statistics depend on the quantity

$$\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}$$

# The Economics of the Time Series Tests

- GRS show that this quantity has a very insightful representation.
- Let us provide the steps.

# Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ .

- Consider an investment universe that consists of  $N + 1$  assets - the  $N$  test assets as well as the market portfolio.
- The expected return vector of the  $N + 1$  assets is given by

$$\hat{\lambda}_{(N+1) \times 1} = \begin{bmatrix} \hat{\mu}_m^e & \hat{\mu}^{e'} \end{bmatrix}'$$

$1 \times 1$     $1 \times N$

where  $\hat{\mu}_m^e$  is the estimated expected excess return on the market portfolio and  $\hat{\mu}^e$  is the estimated expected excess return on the  $N$  test assets.

## Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ .

► The variance covariance matrix of the  $N + 1$  assets is given by

$${}_{(N+1) \times (N+1)} \hat{\Phi} = \begin{bmatrix} \hat{\sigma}_m^2 & \hat{\beta}' \hat{\sigma}_m^2 \\ \hat{\beta} \hat{\sigma}_m^2 & \hat{V} \end{bmatrix}$$

where

$\hat{\sigma}_m^2$  is the estimated variance of the market factor.

$\hat{\beta}$  is the  $N$ -vector of market loadings and  $\hat{V}$  is the covariance matrix of the  $N$  test assets.

## Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ .

- Notice that the covariance matrix of the  $N$  test assets is

$$\hat{V} = \hat{\beta}\hat{\beta}'\hat{\sigma}_m^2 + \hat{\Sigma}$$

- The squared tangency portfolio of the  $N + 1$  assets is

$$SR_{TP}^2 = \hat{\lambda}'\hat{\Phi}^{-1}\hat{\lambda}$$

# Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ .

► Notice also that the inverse of the covariance matrix is

$$\hat{\Phi}^{-1} = \begin{bmatrix} (\hat{\sigma}_m^2)^{-1} + \hat{\beta}'\hat{\Sigma}^{-1}\hat{\beta}, & -\hat{\beta}'\hat{\Sigma}^{-1} \\ -\hat{\Sigma}^{-1}\hat{\beta}, & \hat{\Sigma}^{-1} \end{bmatrix}$$

► Thus, the squared Sharpe ratio of the tangency portfolio could be represented as



# Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ .

$$S\hat{R}_{TP}^2 = \left(\frac{\hat{\mu}_m^e}{\hat{\sigma}_m}\right)^2 + \left[ (\hat{\mu}^e - \hat{\beta}\hat{\mu}_m^e)' \hat{\Sigma}^{-1} (\hat{\mu}^e - \hat{\beta}\hat{\mu}_m^e) \right]$$

$$S\hat{R}_{TP}^2 = S\hat{R}_m^2 + \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

or

$$\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} = S\hat{R}_{TP}^2 - S\hat{R}_m^2$$

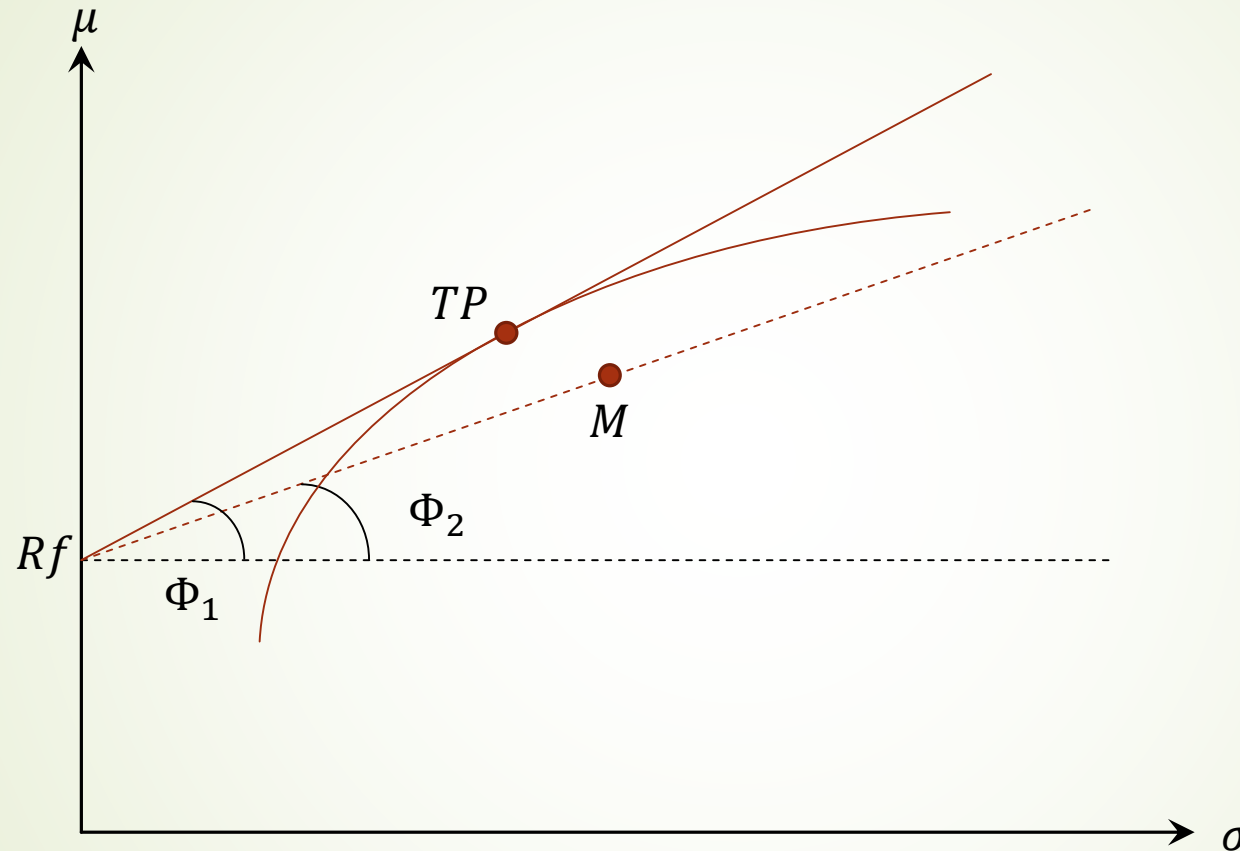
## Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ .

- In words, the  $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$  quantity is the difference between the squared Sharpe ratio based on the  $N + 1$  assets and the squared Sharpe ratio of the market portfolio.
- If the CAPM is correct then these two Sharpe ratios are identical in population, but not identical in sample due to estimation errors.

# Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ .

- The test statistic examines how close the two sample Sharpe ratios are.
- Under the CAPM, the extra  $N$  test assets do not add anything to improving the risk return tradeoff.
- The geometric description of  $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$  is given in the next slide.

# Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ .



$$\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} = \Phi_1^2 - \Phi_2^2$$

# Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ .

► So we can rewrite the previously derived test statistics as

$$J_1 = T \frac{S\hat{R}_{TP}^2 - S\hat{R}_m^2}{1 + S\hat{R}_m^2} \sim \chi^2(N)$$

$$J_3 = \frac{T - N - 1}{N} \times \frac{S\hat{R}_{TP}^2 - S\hat{R}_m^2}{1 + S\hat{R}_m^2} \sim F(N, T - N - 1)$$

# Session #6: Asset Pricing Models with Time Varying Beta

# Asset Pricing Models with Time Varying Beta

- We consider for simplicity only the one factor CAPM – extensions follow the same vein.
- Let us model beta variation with the lagged dividend yield or any other macro variable – again for simplicity we consider only one information, predictive, macro, or lagged variable.



# Asset Pricing Models with Time Varying Beta

Typically, the set of predictive variables contains the dividend yield, the term spread, the default spread, the yield on a T-bill, inflation, lagged market return, market volatility, market illiquidity, etc.

# Conditional Models

Here is a conditional asset pricing specification:

$$r_{it}^e = \alpha_i + \beta_{it}r_{mt}^e + \varepsilon_{it}$$

$$\beta_{it} = \beta_{i0} + \beta_{i1}z_{t-1}$$

$$z_t = a + bz_{t-1} + \eta_t$$

$$E(r_{mt}^e | z_{t-1}) = \mu_m^e$$

$$\text{cov}(\varepsilon_{it}, \eta_t) = 0$$

# Conditional Asset Pricing Models

- Substituting beta back into the asset pricing equation yields.

$$r_{it}^e = \alpha_i + \beta_{i0}r_{mt}^e + \beta_{i1}r_{mt}^e z_{t-1} + \varepsilon_{it}$$

- Interestingly, the one factor conditional CAPM becomes a two factor unconditional model – the first factor is the market portfolio, while the second is the interaction of the market with the lagged variable.

# Conditional Asset Pricing Models

- You can use the statistics  $J_1$  through  $J_4$  to test such models.
- If we have  $K$  factors and  $M$  predictive variables then the  $K$ -conditional factor model becomes a  $K + KM$  - unconditional factor model.
- If you only scale the market beta, as is typically the case, we have an  $M + K$  unconditional factor model.

# Conditional Moments

Suppose you are at time  $t$  – what is the discount factor for time  $t + 2$ ?

$$\begin{aligned}r_{it+2}^e &= \alpha_i + \beta_{i0}r_{mt+2}^e + \beta_{i1}r_{mt+2}^e z_{t+1} + \varepsilon_{it+2} \\ &= \alpha_i + \beta_{i0}r_{mt+2}^e + \beta_{i1}r_{mt+2}^e \times (a + bz_t + \eta_{t+1}) + \varepsilon_{it+2} \\ &= \alpha_i + \beta_{i0}r_{mt+2}^e + a\beta_{i1}r_{mt+2}^e + \beta_{i1}br_{mt+2}^e z_t + \beta_{i1}r_{mt+2}^e \eta_{t+1} + \varepsilon_{it+2}\end{aligned}$$

# Conditional Moments

$$\begin{aligned} E[r_{it+2}^e | z_t] &= \alpha_i + \beta_{i0}\mu_m^e + a\beta_{i1}\mu_m^e + \beta_{i1}b\mu_m^e z_t \\ &= \alpha_i + (\beta_{i0} + a\beta_{i1})\mu_m^e + \beta_{i1}b\mu_m^e z_t \end{aligned}$$

- Notice that here the discount factor, or the conditional expected return, is no longer constant through time.
- Rather, it varies with the macro variable.

# Conditional Moments

Could you derive a general formula – in particular – you are at time  $t$  what is the expected return for time  $t + T$  as a function of the model parameters as well as  $z_t$ ?



# Conditional Moments

- Next, the conditional covariance matrix – the covariance at time  $t + 1$  given  $z_t$  – is given by

$$V[r_{t+1}^e | z_t] = (\beta_0 + \beta_1 z_t)(\beta_0 + \beta_1 z_t)' \sigma_m^2 + \Sigma$$

where  $\beta_0$  and  $\beta_1$  are the  $N$ -asset versions of  $\beta_{i0}$  and  $\beta_{i1}$  respectively.

# Conditional Moments

- Could you derive a general formula – in particular – you are at time  $t$  what is the conditional covariance matrix for time  $t + T$  as a function of the model parameters as well as  $z_t$ ?
- Could you derive general expressions for the conditional moments of *cumulative* return?

# Conditional versus Unconditional Models

There are different ways to model beta variation. Here we used lagged predictive variables; other applications include using firm level variables such as size and book market to scale beta as well as modeling beta as an autoregressive process.

# Conditional versus Unconditional Models

- You can also model time variation in the risk premiums in addition to or instead of beta variations.
- Asset pricing tests show that conditional models typically outperform their unconditional counterparts.

# Different Ways to Model Beta Variation

- The base case: beta is constant, or time invariant.
- Case II: beta varies with macro conditions

$$\beta_{it} = \beta_{i0} + \beta_{i1}z_{t-1}$$

$$z_t = a + bz_{t-1} + \varepsilon_t$$

# Different Ways to Model Beta Variation

- Case III: beta varies with firm-level size and the book-to-market ratio

$$\beta_{it} = \beta_{i0} + \beta_{i1}size_{i,t-1} + \beta_{i2}bm_{i,t-1}$$

- Case IV: beta is some function of both macro and firm-level variables as well as their interactions:

$$\beta_{it} = f(z_{t-1}, size_{i,t-1}, bm_{i,t-1})$$

# Different Ways to Model Beta Variation

- ➔ Case V: beta follows an auto-regressive AR(1) process

$$\beta_{it} = a + b\beta_{i,t-1} + v_{it}$$



# Single vs. Multiple Factors

- Notice that we focus on a single factor single macro variable.
- We can expand the specification to include more factors and more macro and firm-level variables.
- Even if we expand the number of factors it is common to model variation only in the market beta, while the other risk loadings are constant.
- Some scholars model time variations in all factor loadings.

# Testing Conditional Models

- You can implement the  $J_1 - J_4$  test statistics only to those cases where beta is either constant or it varies with macro variables.
- Those specifications involving firm-level characteristics require cross sectional tests.
- The last specification (AR(1)) requires Kalman filtering methods involving state space representations.

# Session #7: GMVP, Tracking Error Volatility, Large scale covariance matrix

# GMVP

- Of particular interest to academics and practitioners is the Global Minimum Volatility Portfolio.
- For two distinct reasons:
  1. No need to estimate the notoriously difficult to estimate  $\mu$ .
  2. Low volatility stocks have been found to outperform high volatility stocks.

# GMVP Optimization

$$\begin{array}{ll} \min & w'Vw \\ \text{s.t} & w'l = 1 \end{array}$$

➤ Solution:  $w_{GMVP} = \frac{V^{-1}l}{l'V^{-1}l}$

- No analytical solution in the presence of portfolio constraints – such as no short selling.

# GMVP Optimization

- Ex ante, the GMVP is the lowest volatility portfolio among all efficient portfolios.
- Ex ante, it is also the lowest mean portfolio, but ex post it performs reasonably well in delivering high payoffs. That is related to the volatility anomaly: low volatility stocks have delivered, on average, higher payoffs than high volatility stocks.

# The Tracking Error Volatility (TEV) Portfolio

- Actively managed funds are often evaluated based on their ability to achieve high return subject to some constraint on their Tracking Error Volatility (TEV).
- In that context, a managed portfolio can be decomposed into both passive and active components.
- TEV is the volatility of the active component.



# The Tracking Error Volatility (TEV) Portfolio

- The passive component is the benchmark portfolio.
- The benchmark portfolio changes with the investment objective.
- For instance, if you invest in S&P500 stocks the proper benchmark would be the S&P index.

# Tracking Error Volatility:

## The Benchmark and Active Portfolios

- ▶ Let  $q$  be the vector of weights of the benchmark portfolio.
- ▶ Then the expected return and variance of the benchmark portfolio are given by

$$\begin{aligned}\mu_B &= q' \mu \\ \sigma_B^2 &= q' V q\end{aligned}$$

where, as usual,  $\mu$  and  $V$  are the vector of expected return and the covariance matrix of stock returns.

# Tracking Error Volatility:

## The Benchmark and Active Portfolios

- The  $V$  matrix can be estimated in different methods – most prominent of which will be discussed here.
- The active fund manager attempts to outperform this benchmark.
- Let  $x$  be the vector of deviations from the benchmark, or the active part of the managed portfolio.
- Of course, the sum of all the components of  $x$ , by construction, must be equal to zero.

# The Mathematics of TEV

- So the fund manager invests  $w = q + x$  in stocks,  $q$  is the passive part of the portfolio and  $x$  is the active part.
- Notice that  $\sigma_{\xi}^2 = x'Vx$  is the tracking error variance.
- Also notice that the expected return and volatility of the chosen portfolio are

$$\begin{aligned}\mu_p &= \mu_B + x'\mu \\ \sigma_p^2 &= w'Vw = \sigma_B^2 + 2q'Vx + \sigma_{\xi}^2\end{aligned}$$

# The Mathematics of TEV

The optimization problem is formulated as

$$\begin{aligned} \max_x \quad & x' \mu \\ \text{s.t.} \quad & x' l = 0 \\ & x' V x = \vartheta \end{aligned}$$

# The Mathematics of TEV

The resulting active part of the portfolio,  $x$ , is given by

$$x = \pm \sqrt{\frac{\vartheta}{e}} V^{-1} \left( \mu - \frac{a}{c} \iota \right)$$

where

$$\mu' V^{-1} \mu = b$$

$$\mu' V^{-1} \iota = a$$

$$\iota' V^{-1} \iota = c$$

$$a - b^2/c = e$$

# Caveats about the Minimum TEV Portfolio

- Richard Roll (distinguished UCLA professor) points out that the solution is independent on the benchmark.
- Put differently, the active part of the portfolio  $x$  is totally independent of the passive part  $q$ .
- Of course, the overall portfolio  $q + x$  is impacted by  $q$ .
- The unexpected result is that the active manager pays no attention to the assigned benchmark. So it does not really matter if the benchmark is S&P or any other index.



# TEV with total Volatility Constraint

(based on Jorion – an Expert in Risk Management)

- Given the drawbacks underlying the TEV portfolio we add one more constraint on the total portfolio volatility.
- The derived active portfolio displays two advantages.
- First, its composition does depend on the benchmark.
- Second, the systematic volatility of the portfolio is controlled by the investor.

# TEV with total volatility constraint

- The optimization is formulated as

$$\max_x x' \mu$$

s. t.

$$x' \iota = 0$$

$$x' V x = \vartheta$$

$$(q + x)' V (q + x) = \sigma_p^2$$

- Home assignment: derive the optimal solution.

# Estimating the Large Scale Covariance Matrix

# Estimating the Covariance Matrix:

- There are various applications in financial economics which use the covariance matrix as an essential input.
- The Global Minimum Variance Portfolio, the minimum tracking error volatility portfolio, the mean variance efficient frontier, and asset pricing tests are good examples.
- In what follows I will present the most prominent estimation methods of the covariance matrix.

# The Sample Covariance Matrix (Denoted $S$ )

- This method uses sample estimates.
- Need to estimate  $N(N + 1)/2$  parameters which is a lot.
- You can use excel to estimate all variances and co-variances which is tedious and inefficient.
- Here is a much more efficient method.
- Consider  $T$  monthly returns on  $N$  risky assets.
- We can display those returns in a  $T$  by  $N$  matrix  $R$ .

# The Sample Covariance Matrix (Denoted $S$ )

- Estimate the mean return of the  $N$  assets – and denote the  $N$ -vector of the mean estimates by  $\hat{\mu}$ .
- Next, compute the deviations of the return observations from their sample means:

$$\hat{E} = R - \iota_T \hat{\mu}'$$

where  $\iota_T$  is a  $T$  vector of ones. Then the sample covariance matrix is estimated as

$$S = \frac{1}{T} \hat{E}' \hat{E}$$

# The Equal Correlation Based

## Covariance Matrix (Denoted F):

- Estimate all  $N(N - 1)/2$  pair-wise correlations between any two securities and take the average.
- Let  $\bar{\rho}$  be that average correlation, let  $s_{ii}$  be the estimated variance of asset  $i$ , and let  $s_{jj}$  be the estimated variance of asset  $j$ , both estimates are the  $i$ -th and  $j$ -th elements of the diagonal of  $S$ .
- Then the matrix  $F$  follows as
$$F(i, i) = s_{ii}$$
$$F(i, j) = \bar{\rho} \sqrt{s_{ii} s_{jj}}$$



# The Factor Based Covariance Matrix:

- Consider the time series regression

$$r_t = \alpha + \beta \times F_t + e_t$$

where  $r_t$  is an  $N$  vector of returns at time  $t$  and  $F_t$  is a set of  $K$  factors. Factor means are denoted by  $\mu_F$

- Notice that the mean return is given by

$$\mu = \alpha + \beta \times \mu_F$$

# The Factor Based Covariance Matrix

- Thus deviations from the means are given by

$$r_t - \mu = \beta \times (F_t - \mu_F) + e_t$$

- The factor based covariance matrix is estimated by

$$\hat{V} = \hat{\beta} \Sigma_{FF} \hat{\beta}' + \hat{\Psi}$$

Here,  $\hat{\beta}$  is an  $N$  by  $K$  matrix of factor loadings and  $\Psi$  is a diagonal matrix with each element represents the idiosyncratic variance of each of the assets.

# The Factor Based Covariance Matrix – Number of Parameters

- This procedure requires the estimation of  $NK$  betas as well as  $K$  variances of the factors,  $K(K - 1)/2$  correlations of those factors, and  $N$  firm specific variances.
- Overall, you need to estimate  $NK + K + K(K - 1)/2 + N$  parameters, which is considerably less than  $N(N + 1)/2$  since  $K$  is much smaller than  $N$ .
- For instance, using a single factor model – the number of parameters to be estimated is only  $2N + 1$ .

# Steps for Estimating the Factor Based Covariance Matrix

1. Run the MULTIVARIATE regression of stock returns on asset pricing factors

$$R_{T \times N} = \iota_{T \times 1} \alpha'_{1 \times N} + F_{T \times K} \beta'_{K \times N} + E_{T \times N} = \check{F}_{T \times (K+1)} \cdot \theta'_{(K+1) \times N} + E_{T \times N}$$

where  $\check{F} = [\iota_T, F]$   
 $\theta = [\alpha, \beta]$

# Steps for Estimating the Factor Based Covariance Matrix

2. Estimate  $\theta$

$$\hat{\theta} = \left[ (\check{F}'\check{F})^{-1} \check{F}'R \right]' = R'\check{F}(\check{F}'\check{F})^{-1} = [\hat{\alpha}, \hat{\beta}]$$

and retain  $\beta$  only.

# Steps for Estimating the Factor Based Covariance Matrix

3. Estimate the covariance matrix of  $E$

$$\hat{\Delta}_{N \times N} = \frac{1}{T} E'_{N \times T} E_{T \times N}$$

4. Let  $\hat{\Psi}$  be a diagonal matrix with the (i,i)-th component being equal to the (i,i)-th component of  $\hat{\Delta}$ .

# Steps for Estimating the Factor Based Covariance Matrix

5. Compute  $V = F - \iota \cdot \hat{\mu}'_f$   
 $T \times K \quad T \times K \quad T \times 1 \quad 1 \times K$

where  $\hat{\mu}_f$  is the mean return of the factors.

6. Estimate:  $\hat{\Sigma}_F = \frac{1}{T} V' \cdot V$   
 $K \times K \quad T \quad K \times T \quad T \times K$



# Steps for Estimating the Factor Based Covariance Matrix

$$7. \quad \hat{V}_{N \times N} = \hat{\beta}_{N \times K} \hat{\Sigma}_f_{K \times K} \hat{\beta}'_{K \times N} + \hat{\Psi}_{N \times N}$$

This is the estimated covariance matrix of stock returns.

8. Notice – there is no need to run  $N$  individual regressions! Use multivariate specifications.

# A Shrinkage Approach –

## Based on a Paper by Ledoit and Wolf (LW)

- There is a well perceived paper (among Wall Street quants) by LW demonstrating an alternative approach to estimating the covariance matrix.
- It had been claimed to deliver superior performance in reducing tracking errors relative to benchmarks as well as producing higher Sharpe ratios.
- Here are the formal details.

# A Shrinkage Approach –

## Based on a Paper by Ledoit and Wolf (LW)

- Let  $S$  be the sample covariance matrix, let  $F$  be the equal correlation based covariance matrix, and let  $\delta$  be the shrinkage intensity.  $S$  and  $F$  were derived earlier.
- The operational shrinkage estimator of the covariance matrix is given by  $\bar{V} = \hat{\delta}^* \times F + (1 - \hat{\delta}^*) \times S$
- Notice that  $F$  is the shrinkage target.

# The Shrinkage Intensity

LW propose the following shrinkage intensity, based on optimization:

$$\hat{\delta}^* = \max \left[ 0, \min \left\{ \frac{k}{T}, 1 \right\} \right]$$

where  $T$  is the sample size and  $k$  is given as  $k = \frac{\pi - \eta}{\gamma}$

and where  $\gamma = \sum_{i=1}^N \sum_{j=1}^N (f_{ij} - s_{ij})^2$  with  $s_{ij}$  being the  $(i, j)$  component of  $S$  and  $f_{ij}$  is the  $(i, j)$  component of  $F$ .

# The Shrinkage Intensity

$$\pi = \sum_{i=1}^N \sum_{j=1}^N \pi_{ij}$$

$$\pi_{ij} = \frac{1}{T} \sum_{t=1}^T \{(r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j) - s_{ij}\}^2$$

$$\eta = \sum_{i=1}^N \pi_{ii} + \sum_{i=1}^N \sum_{j=1, j \neq i}^N \frac{\bar{\rho}}{2} \left( \sqrt{\frac{s_{jj}}{s_{ii}}} \vartheta_{ii,ij} + \sqrt{\frac{s_{ii}}{s_{jj}}} \vartheta_{jj,ij} \right)$$

$$\vartheta_{ii,ij} = \frac{1}{T} \sum_{t=1}^T \{(r_{it} - \bar{r}_i)^2 - s_{ii}\} \{(r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j) - s_{ij}\}$$

$$\vartheta_{jj,ij} = \frac{1}{T} \sum_{t=1}^T \{(r_{jt} - \bar{r}_j)^2 - s_{jj}\} \{(r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j) - s_{ij}\}$$

and with  $r_{it}$  and  $\bar{r}_i$  being the time  $t$  return on asset  $i$  and the time-series average of return on asset  $i$ , respectively.

# The Shrinkage Intensity – A Naïve Method

If you get overwhelmed by the derivation of the shrinkage intensity it would still be useful to use a naïve shrinkage approach, which often even works better. For instance, you can take equal weights:

$$\bar{V} = \frac{1}{2}F + \frac{1}{2}S$$



# Backtesting

- We have proposed several methods for estimating the covariance matrix.
- Which one dominates?
- We can backtest all specifications.
- That is, we can run a “horse race” across the various models searching for the best performer.
- There are two primary methods for backtesting – rolling versus recursive schemes.



# The Rolling Scheme

- You define the first estimation window.
- It is well received to use the first 60 sample observations as the first estimation window.
- Based on those 60 observations derive the GMVP under each of the following methods:

# Competing Covariance Estimates

- The sample based covariance matrix
- The equal correlation based covariance matrix
- Factor model using the market as the only factor
- Factor model using the Fama French three factors
- Factor model using the Fama French plus Momentum factors
- The LW covariance matrix – either the full or the naïve method.

# Out of Sample Returns

- Then given the GMVPs compute the actual returns on each of the derived strategies.
- For instance, if the derived strategy at time  $t$  is  $w_t$  then the realized return at time  $t+1$  would be

$$R_{p,t+1} = w_t' \times R_{t+1}$$

where  $R_{t+1}$  is the realized return at time  $t + 1$  on all the  $N$  investable assets.

# Out of Sample Returns

- Suppose you rebalance every six months – derive the out of sample returns also for the following 5 months

$$R_{p,t+2} = w_t' \times R_{t+2}$$

$$R_{p,t+3} = w_t' \times R_{t+3}$$

$$R_{p,t+4} = w_t' \times R_{t+4}$$

$$R_{p,t+5} = w_t' \times R_{t+5}$$

$$R_{p,t+6} = w_t' \times R_{t+6}$$

- Then at time  $t+6$  you re-derive the GMVPs.

# The Recursive Scheme

- A recursive scheme is using an expanding window.
- That is, you first estimate the GMVPs based on the first 60 observations, then based on 66 observations, and so on, while in the rolling scheme you always use the last 60 observations.
- Pros: the recursive scheme uses more observations.
- Cons: since the covariance matrix may be time varying perhaps you better drop initial observations.

# Out of Sample Returns

- So you generate out of sample returns on each of the strategies starting from time  $t + 1$  till the end of sample, which we typically denote by  $T$ .
- Next, you can analyze the out of sample returns.
- For instance, you can form the table on the next page and examine which specification has been able to deliver the best performance.

# Out of Sample Returns

	Rolling Scheme						Recursive Scheme					
	S	F	MKT	FF	FF+ MOM	LW	S	F	MKT	FF	FF+ MOM	LW
Mean												
STD												
SR												
SP (5%)												
alpha												
IR												



# Out of Sample Returns

In the above table:

- Mean is the simple mean of the out of sample returns
- STD is the volatility of those returns
- SR is the associated Sharpe ratio obtained by dividing the difference between the mean return and the mean risk free rate by STD.
- SP is the shortfall probability with a 5% threshold applied to the monthly returns.

# Out of Sample Returns

In the above table:

- ▶ alpha is the intercept in the regression of out of sample EXCESS returns on the contemporaneous market factor (market return minus the risk-free rate).
- ▶ IR is the information ratio – obtained by dividing alpha by the standard deviation of the regression error, not the STD above.
- ▶ Of course, higher SR, higher alpha, higher IR are associated with better performance.

# High Frequency Data

- ▶ Let  $r_t$  denote the continuously compounded one-year return for year  $t$ , where  $r_t$  are independent variables from  $N(\mu, \sigma^2)$  (iid).
- ▶ Based on  $T$  annual observations, we can estimate  $\mu$  and  $\sigma^2$ :

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t, \text{ and } \hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=1}^T (r_t - \hat{\mu})^2.$$

- ▶ We have  $\widehat{Var}(\hat{\mu}) = \frac{\hat{\sigma}^2}{T}$  and  $\widehat{Var}(\hat{\sigma}^2) = \frac{2\hat{\sigma}^4}{T}$ .

# High Frequency Data – continued

- Now, partition each year to  $N$  equally-separated intervals (e.g., 5 minutes), so that we have  $N$  returns for each year  $t$ :

$$r_{t,1}, r_{t,2}, \dots, r_{t,N} \text{ so that } r_t = \sum_{n=1}^N r_{t,n}.$$

- Overall, we have  $T \cdot N$  observations, and denote:  $\mu_n = E[r_{t,n}]$  (the expected return of one interval).

- Using the iid property, we can estimate:  $\hat{\mu}_n = \frac{\hat{\mu}}{N}$  and  $\hat{\sigma}_n^2 = \frac{\hat{\sigma}^2}{N}$ .

- And we have:  $Var(\hat{\mu}) = N^2 Var(\hat{\mu}_n)$  and  $Var(\hat{\sigma}^2) = N^2 Var(\hat{\sigma}_n^2)$ .

# High Frequency Data – continued

## – The variance of $\hat{\mu}$ and of $\hat{\sigma}^2$

- As usual,  $Var(\hat{\mu}_n) = \frac{\hat{\sigma}_n^2}{TN}$
- Therefore we have (also using the results from previous slide):  
$$Var(\hat{\mu}) = N^2 Var(\hat{\mu}_n) = \frac{N^2 \hat{\sigma}_n^2}{TN} = \frac{N \hat{\sigma}_n^2}{T} = \frac{\hat{\sigma}^2}{T}.$$
- The conclusion regarding  $\hat{\mu}$ : its variance is the same whether it is estimated based on  $T$  annual observations or whether it is estimated using  $T \cdot N$  high-frequency observations.  
There is no gain in using high-frequency observations.
- We will see a different result for  $\hat{\sigma}^2$ .

# High Frequency Data – continued

## – The variance of $\hat{\mu}$ and of $\hat{\sigma}^2$

- What is the variance of  $\hat{\sigma}^2$  based on  $T \cdot N$  high-frequency observations?
- As usual,  $Var(\hat{\sigma}_n^2) = \frac{2\hat{\sigma}_n^4}{TN}$ .
- $Var(\hat{\sigma}^2) = N^2 Var(\hat{\sigma}_n^2) = \frac{2N^2\hat{\sigma}_n^4}{TN} = \frac{1}{N} \cdot \frac{2\hat{\sigma}^4}{T}$ .
- The variance of  $\hat{\sigma}^2$  is much smaller with the partitioning !
- Moreover, the larger the partition is (i.e.,  $N$  is larger), the smaller the variance of the variance is.



# Session #8: Principal Component Analysis (PCA)



# PCA

- The aim is to extract  $K$  common factors to summarize the information of a panel of rank  $N$ .
- In particular, we have a  $T \times N$  panel of stock returns where  $T$  is the time dimension and  $N (< T)$  is the number of firms – of course  $K \ll N$

$$\mathbf{R}_{T \times N} = [R_1, \dots, R_N]$$

- The PCA is an operation on the sample covariance matrix of stock returns.

# PCA

► You have returns on  $N$  stocks for  $T$  periods

$$\begin{matrix} R_1, \dots, R_N \\ T \times 1 \quad T \times 1 \end{matrix}$$

$$\tilde{R}_1 = R_1 - \hat{\mu}_1 \dots \tilde{R}_N = R_N - \hat{\mu}_N$$

let

$$\tilde{R} = [\tilde{R}_1, \tilde{R}_2, \dots, \tilde{R}_N]$$

$$\hat{V} = \frac{1}{T} \tilde{R}' \tilde{R}$$

# PCA

- Extract  $K$ -Eigen vectors corresponding to the largest  $K$ -Eigen values.
- Each of the Eigen vector is an  $N$  by 1 vector.
- The extraction mechanism is as follows.
- The first Eigen vector is obtained as

$$\begin{aligned} \max_{w_1} \quad & w_1' \hat{V} w_1 \\ \text{s.t.} \quad & w_1' w_1 = 1 \end{aligned}$$

# PCA

►  $w_1$  is an Eigen vector since

$$\hat{V}w_1 = \lambda_1 w_1$$

Moreover

$$\lambda_1 = w_1' \hat{V} w_1$$

►  $\lambda_1$  is therefore the highest Eigen value.

# PCA

➔ Extracting the second Eigen vector

$$\begin{array}{ll} \max_{w_2} & w_2' \hat{V} w_2 \\ \text{s. t} & w_2' w_2 = 1 \\ & w_1' w_2 = 0 \end{array}$$

# PCA

- The optimization yields:

$$\hat{V}w_2 = \lambda_2 w_2$$

$$\lambda_2 = w_2' \hat{V} w_2 < \lambda_1$$

- The second Eigen value is smaller than the first due to the presence of one extra constraint in the optimization – the orthogonality constraint.

# PCA

► The  $K$ -th Eigen vector is derived as

$$\begin{aligned} \max \quad & w_K' \hat{V} w_K \\ \text{s. t} \quad & w_K' w_K = 1 \\ & w_K' w_1 = 0 \\ & w_K' w_2 = 0 \\ & \vdots \\ & w_K' w_{K-1} = 0 \end{aligned}$$



# PCA

► The optimization yields:

$$\hat{V}w_K = \lambda_K w_K$$

$$\lambda_K = w_K' \hat{V} w_K < \lambda_{K-1} < \dots < \lambda_2 < \lambda_1$$

# PCA

- Then - each of the  $K$ -Eigen vectors delivers a unique asset pricing factor.
- Simply, multiply excess stock returns by the Eigen vectors:

$$\begin{matrix} F_1 \\ T \times 1 \end{matrix} = \begin{matrix} R^e \\ T \times N \end{matrix} \cdot \begin{matrix} w_1 \\ N \times 1 \end{matrix}$$

⋮

$$\begin{matrix} F_K \\ T \times 1 \end{matrix} = \begin{matrix} R^e \\ T \times N \end{matrix} \cdot \begin{matrix} w_K \\ N \times 1 \end{matrix}$$

# PCA

- Recall, the basic idea here is to replace the original set of  $N$  variables with a lower dimensional set of  $K$ -factors ( $K \ll N$ ).
- The contribution of the  $j$ -th Eigen vector to explain the covariance matrix of stock returns is

$$\frac{\lambda_j}{\sum_{i=1}^N \lambda_i}$$

# PCA

- Typically the first three Eigen vectors explain over and above 95% of the covariance matrix.
- What does it mean to “*explain the covariance matrix*”?  
Coming up soon!

# Understanding the PCA: Digging Deeper

➔ The covariance matrix can be decomposed as

$$\hat{V} = [w_1, \dots, w_N] \begin{bmatrix} \lambda_1, \dots, 0 \\ \vdots & \ddots \\ 0, \dots, \lambda_N \end{bmatrix} \begin{bmatrix} w_1' \\ \vdots \\ w_N' \end{bmatrix}$$

# Understanding the PCA: Digging Deeper

- If some of the  $\lambda - s$  are either zero or negative – the covariance matrix is not properly defined -- it is not positive definite.
- In fixed income analysis – there are three prominent Eigen vectors, or three factors.
- The first factor stands for the term structure level, the second for the term structure slope, and the third for the curvature of the term structure.

# Understanding the PCA: Digging Deeper

- In equity analysis, the first few (up to three) principal components are prominent.
- Others are around zero.
- The attempt is to replace the sample covariance matrix by the matrix  $\tilde{V}$  which mostly summarizes the information in the sample covariance matrix.



# Understanding the PCA: Digging Deeper

► The matrix  $\tilde{V}$  is given by

$$\tilde{V}_{N \times N} = [w_1, \dots, w_k] \begin{bmatrix} \lambda_1, \dots, 0 \\ \vdots & \ddots \\ 0, \dots, \lambda_k \end{bmatrix} \begin{bmatrix} w_1' \\ \vdots \\ w_k' \end{bmatrix}$$

► Of course, the dimension of  $\tilde{V}$  is  $N$  by  $N$ .

► However, its rank is  $K$ , thus the matrix is not invertible.

## Is $\hat{V}$ close to $\tilde{V}$ ?

- This is the same as asking: what does it mean to *explain the sample covariance matrix*?

# Is $\hat{V}$ close to $\tilde{V}$ ?

► Let us represent returns as

$$r_{1t} = \alpha_1 + \underbrace{\beta_{11}f_{1t} + \beta_{12}f_{2t} + \cdots + \beta_{1K}f_{Kt}}_{\tilde{r}_{1t}} + \varepsilon_{1t} \quad \forall t = 1, \dots, T$$

⋮

$$r_{Nt} = \alpha_N + \underbrace{\beta_{N1}f_{1t} + \beta_{N2}f_{2t} + \cdots + \beta_{NK}f_{Kt}}_{\tilde{r}_{Nt}} + \varepsilon_{Nt} \quad \forall t = 1, \dots, T$$

## Is $\hat{V}$ close to $\tilde{V}$ ?

- ➔ where  $f_{1t} \dots f_{Kt}$  are the principal component based factors and  $\beta_{i1} \dots \beta_{iK}$  are the exposures of firm  $i$  to those factors.

## Is $\hat{V}$ close to $\tilde{V}$ ?

➔  $\hat{V}$  is close enough to  $\tilde{V}$  - if

1. The variances of the residuals cannot be dramatically reduced by adding more factors.
2. The pairwise cross-section correlations of the residuals cannot be considerably reduced by adding more factors.

# The Contribution of the PC-S to Explain Portfolio Variation

► Let  $\beta_i = [\beta_{i1}, \dots, \beta_{iK}]'$  be the exposures of firm  $i$  to the  $K$  common factors.

► Let  $w_p$  be an  $N$ -vector of portfolio weights:

$$w_p = [w_{1p}, w_{2p}, \dots, w_{Np}]'$$

► Recall,  $R$  is a  $T \times N$  matrix of stock returns.

# The Contribution of the PC-S to Explain Portfolio Variation

- The portfolio's rate of return is

$$R_p = R \cdot w_p$$

$T \times 1 \quad T \times N \quad N \times 1$

- Moreover, the portfolio time t return is given by

$$R_{pt} = w_{1p}r_{1t} + w_{2p}r_{2t} + \dots + w_{Np}r_{Nt} \equiv w_p' \cdot r_t$$

$1 \times N \quad N \times 1$





# The Contribution of the PC-S to Explain Portfolio Variation

➔ Thus,  $\tilde{R}_{pt} = \delta_1 f_{1t} + \delta_2 f_{2t} + \dots + \delta_K f_{Kt}$

where

$$\delta_1 = \underset{1 \times N}{w_p'} \cdot \underset{N \times 1}{\beta_1}$$

$$\delta_2 = \underset{1 \times N}{w_p'} \cdot \underset{N \times 1}{\beta_2}$$

⋮

$$\delta_K = \underset{1 \times N}{w_p'} \cdot \underset{N \times 1}{\beta_K}$$

# The Contribution of the PC-S to Explain Portfolio Variation

Notice that

$\delta_1 \dots \delta_K$  are the  $K$  loadings on the common factors,  
or they are the risk exposures, while  $f_{1t} \dots f_{Kt}$   
are the  $K$ -realizations of the factors at time  $t$ .

# Explain the Portfolio Variance

- The actual variation is

$$\sigma^2(R_{pt}) = w_p' \hat{V} w_p$$

- The approximated variation is

$$\sigma^2(\tilde{R}_{pt}) = \delta_1^2 \text{var}(f_{1t}) + \delta_2^2 \text{var}(f_{2t}) + \dots + \delta_K^2 \text{var}(f_{Kt})$$

- Both quantities are quite similar.

# Explain the Portfolio Variance

- ➔ The contribution of the  $i$ -th PC to the overall portfolios variance is:

$$\frac{\delta_i^2 \text{var} (f_i)}{\sum_{j=1}^k \delta_j^2 \text{var} (f_j)}$$

## Asy. PCA: What if $N > T$ ?

- Then create a  $T \times T$  matrix  $\hat{V}$  and extract  $K$  Eigen vectors – those Eigen vectors are the factors

$$\hat{V} = \frac{1}{N} \hat{E}' \hat{E}$$

where

$$\hat{E}_{T \times N} = R_{T \times N} - \hat{\mu}_{T \times 1} \cdot \iota'_{1 \times N}$$

and  $\hat{\mu}$  is the  $T$ -vector of cross sectional (across- stocks) mean of returns.

# Other Applications

- PCA can be implemented in a host of other applications.
- For instance, you want to predict economic growth with many predictors, say  $M$  where  $M$  is large.
- You have a panel of  $T \times M$  predictors, where  $T$  is the time-series dimension.



# Other Applications

► In a matrix form

$$\mathbf{Z}_{T \times M} = \begin{bmatrix} Z_{11}, Z_{12}, \dots, Z_{1M} \\ \vdots \\ Z_{T1}, Z_{T2}, \dots, Z_{TM} \end{bmatrix}$$

where  $Z_{tm}$  is the  $m$ -th predictor realized at time  $t$ .

# Other Applications

- If  $T > M$  compute the covariance matrix of  $Z$  – then extract  $K$  principal components such that you summarize the  $M$ -dimension of the predictors with a smaller subspace of order  $K \ll M$ .
- You extract  $w_1, \dots, w_K$  Eigen vectors.  
 $M \times 1$        $M \times 1$

# Other Applications

► Then you construct  $K$  predictors:

$$\begin{matrix} Z_1 & = & Z & \cdot & w_1 \\ T \times 1 & & T \times M & & M \times 1 \end{matrix}$$

⋮

$$\begin{matrix} Z_K & = & Z & \cdot & w_K \\ T \times 1 & & T \times M & & M \times 1 \end{matrix}$$

## What if $M > T$ ?

- If  $M > T$  then you extract  $K$  Eigen vectors from the  $T \times T$  matrix.
- In this case, the  $K$ -predictors are the extracted  $K$  Eigen vectors.
- Be careful of a look-ahead bias in real time prediction.

# The Number of Factors in PCA

- An open question is: how many factors/Eigen vectors should be extracted?
- Here is a good mechanism: set  $K_{\max}$  - the highest number of factors.
- Run the following multivariate regression for

$$K = 1, 2, \dots, K_{\max}$$

$$R_{T \times N} = \iota_{T \times 1} + \alpha'_{1 \times N} + F_{T \times K} \beta'_{K \times N} + E_{T \times N}$$

# The Number of Factors in PCA

- ➔ Estimate first the residual covariance matrix and then the average of residual variances:

$$\hat{V}_{N \times N} = \frac{1}{T - K - 1} \hat{E}' \hat{E}$$

$$\hat{\sigma}^2(K) = \frac{tr(\hat{V})}{N}$$

where  $tr(\hat{V})$  is the sum of diagonal elements in  $\hat{V}$

# The Number of Factors in PCA

- Compute for each chosen  $K$

$$PC(K) = \hat{\sigma}^2(K) + K \hat{\sigma}^2(K_{\max}) \frac{N + T}{NT} \cdot \ln \left( \frac{NT}{N + T} \right)$$

and pick  $K$  which minimizes this criterion.

*Notice that with more factors the first component (goodness of fit) diminishes but the second component (penalty factor) rises. There is a tradeoff here.*



# Session #9: Bayesian Econometrics

# Bayes Rule

- ▶ Let  $x$  and  $y$  be two random variables
- ▶ Let  $P(x)$  and  $P(y)$  be the two *marginal* probability distribution functions of  $x$  and  $y$
- ▶ Let  $P(x|y)$  and  $P(y|x)$  denote the corresponding *conditional* pdfs
- ▶ Let  $P(x, y)$  denote the *joint* pdf of  $x$  and  $y$
- ▶ It is known from the law of total probability that the joint pdf can be decomposed as

$$P(x, y) = P(x)P(y|x) = P(y)P(x|y)$$

- ▶ Therefore

$$\begin{aligned} P(y|x) &= \frac{P(y)P(x|y)}{P(x)} \\ &= cP(y)p(x|y) \end{aligned}$$

where  $c$  is the constant of integration (see next page)

- ▶ The Bayes Rule is described by the following proportion

$$P(y|x) \propto P(y)P(x|y)$$

# Bayes Rule

- ▶ Notice that the right hand side retains only factors related to  $y$ , thereby excluding  $P(x)$
- ▶  $P(x)$ , termed the marginal likelihood function, is

$$\begin{aligned}P(x) &= \int P(y)P(x|y)dy \\ &= \int P(x, y)dy\end{aligned}$$

as the conditional distribution  $P(y|x)$  integrates to unity.

- ▶ The marginal likelihood  $P(x)$  is an essential ingredient in computing the model posterior probability, or the probability that a model is correct.
- ▶ The marginal likelihood obtains by integrating out  $y$  from the joint density  $P(x, y)$
- ▶ Similarly, if the joint distribution is  $P(x, y, z)$  and the pdf of interest is  $P(x, y)$  one integrates  $P(x, y, z)$  with respect to  $z$ .

# Bayes Rule

- ▶ The essence of Bayesian econometrics is the Bayes Rule.
- ▶ Ingredients of Bayesian econometrics are parameters underlying a given model, the sample data, the prior density of the parameters, the likelihood function describing the data, and the posterior distribution of the parameters. A predictive distribution is often involved.
- ▶ Indeed, in the Bayesian setup parameters are stochastic while in the classical (non Bayesian) approach parameters are unknown constants.
- ▶ Decision making is based on the posterior distribution of the parameters or the predictive distribution of the data as described below.
- ▶ On the basis of the Bayes rule, in what follows,  $y$  stands for unknown stochastic parameters,  $x$  for the data,  $P(y|x)$  for the posterior distribution,  $P(y)$  for the prior, and  $P(x|y)$  for the likelihood
- ▶ Then the Bayes rule describes the relation between the prior, the likelihood, and the posterior

$$P(y|x) \propto P(y)P(x|y)$$

- ▶ Zellner (1971) is an excellent reference

# Bayes Econometrics in Financial Economics

- ▶ You observe the returns on the market index over  $T$  months:  $r_1, \dots, r_T$
- ▶ Let  $R: [r_1, \dots, r_T]'$  denote the  $T \times 1$  vector of all return realizations
- ▶ Assume that  $r_t \sim N(\mu, \sigma_0^2)$  for  $t = 1, \dots, T$

where

$\mu$  is a stochastic random variable denoting the mean return

$\sigma_0^2$  is the volatility which, at this stage, is assumed to be a known constant

and returns are IID (independently and identically distributed) through time.

- ▶ By Bayes rule

$$P(\mu|R, \sigma_0^2) \propto P(\mu)P(R|\mu, \sigma_0^2)$$

where

$P(\mu|R, \sigma_0^2)$  is the posterior distribution of  $\mu$

$P(\mu)$  is the prior distribution of  $\mu$

and  $P(R|\mu, \sigma_0^2)$  is the joint likelihood of all return realizations.

# Bayes Econometrics: Likelihood

- The likelihood function of a normally distributed return realization is given by

$$P(r_t|\mu, \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2}(r_t - \mu)^2\right)$$

- Since returns are assumed to be IID, the joint likelihood of all realized returns is

$$P(R|\mu, \sigma_0^2) = (2\pi\sigma_0^2)^{-\frac{T}{2}} \exp\left(-\frac{1}{2\sigma_0^2} \sum_{t=1}^T (r_t - \mu)^2\right)$$

- Notice:

$$\begin{aligned} \sum (r_t - \mu)^2 &= \sum [(r_t - \hat{\mu}) + (\hat{\mu} - \mu)]^2 \\ &= vs^2 + T(\mu - \hat{\mu})^2 \end{aligned}$$

since the cross product is zero, and

$$\begin{aligned} v &= T - 1 \\ s^2 &= \frac{1}{T - 1} \sum (r_t - \hat{\mu})^2 \end{aligned}$$

$$\hat{\mu} = \frac{1}{T} \sum r_t$$



# Prior

- The prior is specified by the researcher based on economic theory, past experience, past data, current similar data, etc. Often, the prior is diffuse or non-informative
- For the next illustration, it is assumed that  $P(\mu) \propto c$ , that is, the prior is diffuse, non-informative, in that it *apparently* conveys no information on the parameters of interest.
- I emphasize “apparently” since innocent diffuse priors could exert substantial amount of information about quantities of interest which are non-linear functions of the parameters.
- Informative priors with sound economic appeal are well perceived in financial economics.
- For instance, Kandel and Stambaugh (1996), who study asset allocation when stock returns are predictable, entertain informative prior beliefs weighted against predictability. Pastor and Stambaugh (1999) introduce prior beliefs about expected stock returns which consider factor model restrictions. Avramov, Cederburg, and Kvasnakova (2015) study prior beliefs about predictive regression parameters which are centered around values implied by either prospect theory or the long run risk model of Bansal and Yaron (2004).
- Computing posterior probabilities of competing models (e.g., Avramov (2002)) necessitates the use of informative priors. Here, diffuse priors won't do the work.



# The Posterior Distribution

- ▶ With diffuse prior and normal likelihood, the posterior is

$$P(\mu|R, \sigma_0^2) \propto \exp\left(-\frac{1}{2\sigma_0^2} [vs^2 + T(\mu - \hat{\mu})]^2\right)$$
$$\propto \exp\left(-\frac{T}{2\sigma_0^2} (\mu - \hat{\mu})^2\right)$$

- ▶ The bottom relation follows since only factors related to  $\mu$  are retained
- ▶ The posterior distribution of the mean return is given by

$$\mu|R, \sigma_0^2 \sim N\left(\hat{\mu}, \sigma_0^2/T\right)$$

- ▶ In classical econometrics:

$$\hat{\mu}|R, \sigma_0^2 \sim N\left(\mu, \sigma_0^2/T\right)$$

- ▶ That is, in classical econometrics, the sample estimate of  $\mu$  is stochastic while  $\mu$  itself is an unknown constant.

# Informative Prior

- The prior on the mean return is often modeled as

$$P(\mu) \propto (\sigma_a)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_a^2}(\mu - \mu_a)^2\right)$$

where  $\mu_a$  and  $\sigma_a$  are prior parameters to be specified by the researcher

- The posterior obtains by combining the prior and the likelihood:

$$\begin{aligned} P(\mu|R, \sigma_0^2) &\propto P(\mu)P(R|\mu, \sigma_0^2) \\ &\propto \exp\left[-\frac{1}{2} \frac{(\mu - \mu_a)^2}{\sigma_a^2} + \frac{T}{\sigma_0^2} (\mu - \hat{\mu})^2\right] \\ &\propto \exp\left(-\frac{1}{2} \frac{(\mu - \tilde{\mu})^2}{\tilde{\sigma}^2}\right) \end{aligned}$$

- The bottom relation obtains by **completing the square** on  $\mu$
- Notice, in particular,

$$\begin{aligned} \frac{\mu^2}{\sigma_a^2} + \frac{T}{\sigma_0^2} \mu^2 &= \frac{\mu^2}{\tilde{\sigma}^2} \\ \left(\frac{1}{\sigma_a^2} + \frac{T}{\sigma_0^2}\right) &= \frac{1}{\tilde{\sigma}^2} \end{aligned}$$

# The Posterior Mean

- ▶ Hence, the posterior variance of the mean is

$$\begin{aligned}\tilde{\sigma}^2 &= \left[ \frac{1}{\sigma_a^2} + \frac{1}{\sigma_0^2/T} \right]^{-1} \\ &= (\text{prior precision} + \text{likelihood precision})^{-1}\end{aligned}$$

- ▶ Similarly, the posterior mean of  $\mu$  is

$$\begin{aligned}\tilde{\mu} &= \tilde{\sigma}^2 \left[ \frac{\mu_0}{\sigma_a^2} + \frac{T\hat{\mu}}{\sigma_0^2} \right] \\ &= w_1\mu_0 + w_2\hat{\mu}\end{aligned}$$

where

$$w_1 = \frac{\frac{1}{\sigma_a^2}}{\frac{1}{\sigma_a^2} + \frac{1}{\sigma_0^2/T}} = \frac{\text{prior precision}}{\text{prior precision} + \text{likelihood precision}}$$
$$w_2 = 1 - w_1$$

- ▶ So the posterior mean of  $\mu$  is the weighted average of the prior mean and the sample mean with weights depending on the prior and likelihood precisions, respectively.

# What if $\sigma$ is unknown? – The case of Diffuse Prior

► Bayes:  $P(\mu, \sigma | R) \propto P(\mu, \sigma)P(R | \mu, \sigma)$

► The non-informative prior is typically modeled as

$$P(\mu, \sigma) \propto P(\mu)P(\sigma)$$

$$P(\mu) \propto c$$

$$P(\sigma) \propto \sigma^{-1}$$

► Thus, the joint posterior of  $\mu$  and  $\sigma$  is

$$P(\mu, \sigma | R) \propto \sigma^{-(T+1)} \exp\left(-\frac{1}{2\sigma^2} [vs^2 + T(\mu - \hat{\mu})^2]\right)$$

► The conditional distribution of the mean follows straightforwardly

$$P(\mu | \sigma, R) \text{ is } N\left(\hat{\mu}, \sigma^2 / T\right)$$

► More challenging is to uncover the marginal distributions, which are obtained as

$$P(\mu | R) = \int P(\mu, \sigma | R) d\sigma$$

$$P(\sigma | R) = \int P(\mu, \sigma | R) d\mu$$

# Solving the Integrals: Posterior of $\mu$

► Let  $\alpha = vs^2 + T(\mu - \hat{\mu})^2$

► Then,

$$P(\mu|R) \propto \int_{\sigma=0}^{\infty} \sigma^{-(T+1)} \exp\left(-\frac{\alpha}{2\sigma^2}\right) d\sigma$$

► We do a change of variable

$$\begin{aligned} x &= \frac{\alpha}{2\sigma^2} \\ \frac{d\sigma}{dx} &= -2^{-1} \frac{1}{2} \alpha^{\frac{1}{2}} x^{-1} \frac{1}{2} \\ \sigma^{-T+1} &= \left(\frac{\alpha}{2x}\right)^{-\frac{T+1}{2}} \end{aligned}$$

► Then

$$P(\mu|R) \propto 2^{\frac{T-2}{2}} \alpha^{-\frac{T}{2}} \int_{x=0}^{\infty} x^{\frac{T}{2}-1} \exp(-x) dx$$

Notice

$$\int_{x=0}^{\infty} x^{\frac{T}{2}-1} \exp(-x) dx = \Gamma\left(\frac{T}{2}\right)$$

► Therefore,

$$\begin{aligned} P(\mu|R) &\propto 2^{\frac{T-2}{2}} \Gamma\left(\frac{T}{2}\right) \alpha^{-\frac{T}{2}} \\ &\propto [vs^2 + T(\mu - \hat{\mu})^2]^{-\frac{v+1}{2}} \end{aligned}$$

► We get  $t = \frac{\mu - \hat{\mu}}{s/\sqrt{T}} \sim t(v)$ , corresponding to the Student t distribution with  $v$  degrees of freedom.

# The Marginal Posterior of $\sigma$

- ▶ The posterior on  $\sigma$

$$\begin{aligned} P(\sigma|R) &\propto \int \sigma^{-(T+1)} \exp\left(-\frac{1}{2\sigma^2} [vs^2 + T(\mu - \hat{\mu})]^2\right) d\mu \\ &\propto \sigma^{-(T+1)} \exp\left(-\frac{vs^2}{2\sigma^2}\right) \int \exp\left(-\frac{T}{2\sigma^2} (\mu - \hat{\mu})^2\right) d\mu \end{aligned}$$

- ▶ Let  $z = \frac{\sqrt{T}(\mu - \hat{\mu})}{\sigma}$ , then

$$\frac{dz}{d\mu} = \sqrt{T} \frac{1}{\sigma}$$

- ▶

$$\begin{aligned} P(\sigma|R) &\propto \sigma^{-T} \exp\left(-\frac{vs^2}{2\sigma^2}\right) \int \exp(-z^2/2) dz \\ &\propto \sigma^{-T} \exp\left(-\frac{vs^2}{2\sigma^2}\right) \\ &\propto \sigma^{-(v+1)} \exp\left(-\frac{vs^2}{2\sigma^2}\right) \end{aligned}$$

which corresponds to the inverted gamma distribution with  $v$  degrees of freedom and parameter  $s$

- ▶ The explicit form (with constant of integration) of the inverted gamma is given by

$$P(\sigma|v, s) = \frac{2}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{vs^2}{2}\right)^{v/2} \sigma^{-(v+1)} \exp\left(-\frac{vs^2}{2\sigma^2}\right)$$



# The Multiple Regression Model

- ▶ The regression model is given by

$$y = X\beta + u$$

where

$y$  is a  $T \times 1$  vector of the dependent variables

$X$  is a  $T \times M$  matrix with the first column being a  $T \times 1$  vector of ones

$\beta$  is an  $M \times 1$  vector containing the intercept and  $M-1$  slope coefficients

and  $u$  is a  $T \times 1$  vector of residuals.

- ▶ We assume that  $u_t \sim N(0, \sigma^2) \forall t = 1, \dots, T$  and IID through time
- ▶ The likelihood function is

$$\begin{aligned} P(y|X, \beta, \sigma) &\propto \sigma^{-T} \exp\left(-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta)\right) \\ &\propto \sigma^{-T} \exp\left\{-\frac{1}{2\sigma^2} [vs^2 + (\beta - \hat{\beta})' X'X(\beta - \hat{\beta})]\right\} \end{aligned}$$



# The Multiple Regression Model

where

$$v = T - M$$

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$s^2 = \frac{1}{v} (y - X\hat{\beta})' (y - X\hat{\beta})$$

► It follows since

$$\begin{aligned} (y - X\beta)'(y - X\beta) &= [y - X\hat{\beta} - X(\beta - \hat{\beta})]' [y - X\hat{\beta} - X(\beta - \hat{\beta})] \\ &= (y - X\hat{\beta})'(y - X\hat{\beta}) + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \end{aligned}$$

while the cross product is zero

# Assuming Diffuse Prior

- ▶ The prior is modeled as

$$P(\beta, \sigma) \propto \frac{1}{\sigma}$$

- ▶ Then the joint posterior of  $\beta$  and  $\sigma$  is

$$P(\beta, \sigma | y, X) \propto \sigma^{-(T+1)} \exp \left\{ -\frac{1}{2\sigma^2} \left[ v s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right] \right\}$$

- ▶ The conditional posterior of  $\beta$  is

$$P(\beta | \sigma, y, X) \propto \exp \left( -\frac{1}{2\sigma^2} (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right)$$

which obeys the multivariable normal distribution

$$N(\hat{\beta}, (X'X)^{-1}\sigma^2)$$

# Assuming Diffuse Prior

- ▶ What about the marginal posterior for  $\beta$  ?

$$P(\beta|y, X) = \int P(\beta, \sigma|y, X) d\sigma$$
$$\propto \left[ \nu s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \right]^{-T/2}$$

which pertains to the multivariate student  $t$  with mean  $\hat{\beta}$  and  $T-M$  degrees of freedom

- ▶ What about the marginal posterior for  $\sigma$ ?

$$P(\sigma|y, X) = \int P(\beta, \sigma|y, X) d\beta$$
$$\propto \sigma^{-(\nu+1)} \exp\left(-\frac{\nu s^2}{2\sigma^2}\right)$$

which stands for the inverted gamma with  $T-M$  degrees of freedom and parameter  $s$

- You can simulate the distribution of  $\beta$  in two steps without solving analytically the integral, drawing first  $\sigma$  from its inverted gamma distribution and then drawing from the conditional of  $\beta$  which is normal as shown earlier.

# Bayesian Updating/Learning

- Suppose the initial sample consists of  $T_1$  observations of  $X_1$  and  $y_1$ .
- Suppose further that the posterior distribution of  $(\beta, \sigma)$  based on those observations is given by:

$$\begin{aligned} P(\beta, \sigma | y_1, X_1) &\propto \sigma^{-(T_1+1)} \exp \left[ -\frac{1}{2\sigma^2} (y_1 - X_1\beta)'(y_1 - X_1\beta) \right] \\ &\propto \sigma^{-(T_1+1)} \exp \left\{ -\frac{1}{2\sigma^2} [v_1 s_1^2 + (\beta - \hat{\beta}_1)' X_1' X_1 (\beta - \hat{\beta}_1)] \right\} \end{aligned}$$

where

$$\begin{aligned} v_1 &= T_1 - M \\ \hat{\beta}_1 &= (X_1' X_1)^{-1} X_1' y_1 \\ v_1 s_1^2 &= (y_1 - X_1 \hat{\beta}_1)' (y_1 - X_1 \hat{\beta}_1) \end{aligned}$$

- You now observe one additional sample  $X_2$  and  $y_2$  of length  $T_2$  observations
- The likelihood based on that sample is

$$P(y_2, X_2 | \beta, \sigma) \propto \sigma^{-T_2} \exp \left[ -\frac{1}{2\sigma^2} (y_2 - X_2\beta)'(y_2 - X_2\beta) \right]$$

# Bayesian Updating/Learning

- Combining the posterior based on the first sample (which becomes the prior for the second sample) and the likelihood based on the second sample yields:

$$P(\beta, \sigma | y_1, y_2, X_1, X_2) \propto \sigma^{-(T_1+T_2+1)} \exp \left\{ -\frac{1}{2\sigma^2} [(y_1 - X_1\beta)'(y_1 - X_1\beta) + (y_2 - X_2\beta)'(y_2 - X_2\beta)] \right\}$$
$$\propto \sigma^{-(T_1+T_2+1)} \exp \left\{ -\frac{1}{2\sigma^2} [vs^2 + (\beta - \tilde{\beta})' \mathfrak{M}(\beta - \tilde{\beta})] \right\}$$

where

$$\mathfrak{M} = X_1'X_1 + X_2'X_2$$
$$\tilde{\beta} = \mathfrak{M}^{-1}(X_1'y_1 + X_2'y_2)$$
$$vs^2 = (y_1 - X_1\tilde{\beta})'(y_1 - X_1\tilde{\beta}) + (y_2 - X_2\tilde{\beta})'(y_2 - X_2\tilde{\beta})$$
$$v = T_1 + T_2 - M$$

- Then the posterior distributions for  $\beta$  and  $\sigma$  follow using steps outlined earlier
- With more observations realized you follow the same updating procedure
- Notice that the same posterior would have been obtained starting with diffuse priors and then observing the two samples jointly  $Y=[y_1', y_2']'$  and  $X = [X_1', X_2']'$ .

# The Black-Litterman way of Estimating Mean Returns

- The well-known BL approach to estimating mean return exhibits some Bayesian appeal.
- It is notoriously difficult to propose good estimates for mean returns.
- The sample means are quite noisy – thus asset pricing models -even if misspecified -could give a good guidance.
- To illustrate, you consider a  $K$ -factor model (factors are portfolio spreads) and you run the time series regression

$$r_t^e = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t} + \cdots + \beta_K f_{Kt} + e_t$$

$N \times 1 \quad N \times 1 \quad N \times 1 \quad N \times 1 \quad N \times 1 \quad N \times 1$



# Estimating Mean Returns

➔ Then the estimated excess mean return is given by

$$\hat{\mu}^e = \hat{\beta}_1 \hat{\mu}_{f_1} + \hat{\beta}_2 \hat{\mu}_{f_2} + \dots + \hat{\beta}_K \hat{\mu}_{f_K}$$

where  $\hat{\beta}_1, \hat{\beta}_2 \dots \hat{\beta}_K$  are the sample estimates of factor loadings, and  $\hat{\mu}_{f_1}, \hat{\mu}_{f_2} \dots \hat{\mu}_{f_K}$  are the sample estimates of the factor mean returns.



# The BL Mean Returns

- The BL approach combines a model (CAPM) with some views, either relative or absolute, about expected returns.
- The BL vector of mean returns is given by

$$\mu_{BL} = \begin{bmatrix} \tau & \Sigma \\ 1 \times 1 & N \times N \end{bmatrix}^{-1} + \begin{bmatrix} P' & \Omega^{-1} & P \\ N \times K & K \times K & K \times N \end{bmatrix}^{-1} \cdot \begin{bmatrix} \tau & \Sigma \\ 1 \times 1 & N \times N \end{bmatrix}^{-1} \begin{bmatrix} \mu^{eq} \\ N \times 1 \end{bmatrix} + \begin{bmatrix} P' & \Omega^{-1} \\ N \times K & K \times K \end{bmatrix} \begin{bmatrix} \mu^v \\ K \times 1 \end{bmatrix}$$

# Understanding the BL Formulation

- ➔ We need to understand the essence of the following parameters, which characterize the mean return vector

$$\Sigma, \mu^{eq}, P, \tau, \Omega, \mu^v$$

- ➔ Starting from the  $\Sigma$  matrix – you can choose any of the specifications derived in the previous meetings – either the sample covariance matrix, or the equal correlation, or an asset pricing based covariance, or you could rely on the LW shrinkage approach – either the complex or the naïve one.

# Constructing Equilibrium Expected Returns

- The  $\mu^{eq}$ , which is the equilibrium vector of expected return, is constructed as follows.
- Generate  $\omega_{MKT}$ , the  $N \times 1$  vector denoting the weights of any of the  $N$  securities in the market portfolio based on market capitalization. Of course, the sum of weights must be unity.
- Then, the price of risk is  $\gamma = \frac{\mu_m - Rf}{\sigma_m^2}$  where  $\mu_m$  and  $\sigma_m^2$  are the expected return and variance of the market portfolio.
- Later, we will justify this choice for the price of risk.

# Constructing Equilibrium Expected Returns

- One could pick a range of values for  $\gamma$  going from 1.5 to 2.5 and examine performance of each choice.
- If you work with monthly observations, then switching to the annual frequency does not change  $\gamma$  as both the numerator and denominator are multiplied by 12.
- Having at hand both  $\omega_{MKT}$  and  $\gamma$ , the equilibrium return vector is given by

$$\mu^{eq} = \gamma \Sigma \omega_{MKT}$$

# Constructing Equilibrium Expected Returns

- This vector is called neutral mean or equilibrium expected return.
- To understand why, notice that if you have a utility function that generates the tangency portfolio of the form

$$w_{TP} = \frac{\Sigma^{-1} \mu^e}{\iota' \Sigma^{-1} \mu^e}$$

then using  $\mu^{eq}$  as the vector of excess returns on the  $N$  assets would deliver  $\omega_{MKT}$  as the tangency portfolio.

# What if you Directly Apply the CAPM?

- The question being – would you get the same vector of equilibrium mean return if you directly use the CAPM?
- Yes, if...



# The CAPM based Expected Returns

► Under the CAPM the vector of excess returns is given by

$$\mu^e_{N \times 1} = \beta_{N \times 1} \mu_m^e$$

$$\beta = \frac{\text{cov}(r^e, r_m^e)}{\sigma_m^2} = \frac{\text{cov}(r^e, (r^e)' w_{MKT})}{\sigma_m^2} = \frac{\sum w_{MKT}}{\sigma_m^2}$$

$$CAPM: \mu^e_{N \times 1} = \frac{\sum w_{MKT}}{\sigma_m^2} \mu_m^e = \gamma \sum w_{MKT}$$



# What if you use Directly the CAPM?

Since  $\mu_m^e = (\mu^e)' w_{MKT}$  and  $r_m^e = (r^e)' w_{MKT}$

$$\text{then } \mu^e = \frac{\mu_m^e}{\sigma_m^2} \sum w_{MKT} = \mu^{eq}$$

# What if you use Directly the CAPM?

- So indeed, if you use (i) the sample covariance matrix, rather than any other specification, as well as (ii)

$$\gamma = \frac{\mu_m - Rf}{\sigma_m^2}$$

then the BL equilibrium expected returns and expected returns based on the CAPM are identical.

# The $P$ Matrix: Absolute Views

- In the BL approach the investor/econometrician forms some views about expected returns as described below.
- $P$  is defined as that matrix which identifies the assets involved in the views.
- To illustrate, consider two "absolute" views only.
- The first view says that stock 3 has an expected return of 5% while the second says that stock 5 will deliver 12%.

# The $P$ Matrix: Absolute Views

- In general the number of views is  $K$ .
- In our case  $K = 2$ .
- Then  $P$  is a  $2 \times N$  matrix.
- The first row is all zero except for the fifth entry which is one.
- Likewise, the second row is all zero except for the fifth entry which is one.

# The $P$ Matrix: Relative Views

- Let us consider now two "relative views".
- Here we could incorporate market anomalies into the BL paradigm.
- Market anomalies are cross sectional patterns in stock returns unexplained by the CAPM.
- Example: price momentum, earnings momentum, value, size, accruals, credit risk, dispersion, and volatility.

# Black-Litterman: Momentum and Value Effects

- Let us focus on price momentum and the value effects.
- Assume that both momentum and value investing outperform.
- The first row of  $P$  corresponds to momentum investing.
- The second row corresponds to value investing.
- Both the first and second rows contain  $N$  elements.

# Winner, Loser, Value, and Growth Stocks

- Winner stocks are the top 10% performers during the past six months.
- Loser stocks are the bottom 10% performers during the past six months.
- Value stocks are 10% of the stocks having the highest book-to-market ratio.
- Growth stocks are 10% of the stocks having the lowest book-to-market ratios.



# Momentum and Value Payoffs

- The momentum payoff is a return spread – return on an equal weighted portfolio of winner stocks minus return on equal weighted portfolio of loser stocks.
- The value payoff is also a return spread – the return differential between equal weighted portfolios of value and growth stocks.

# Back to the $P$ Matrix

- Suppose that the investment universe consists of 100 stocks
- The first row gets the value 0.1 if the corresponding stock is a winner (there are 10 winners in a universe of 100 stocks).
- It gets the value -0.1 if the corresponding stock is a loser (there are 10 losers).
- Otherwise it gets the value zero.
- The same idea applies to value investing.
- Of course, since we have relative views here (e.g., return on winners minus return on losers) then the sum of the first row and the sum of the second row are both zero.

# Back to the $P$ Matrix

- More generally, if  $N$  stocks establish the investment universe and moreover momentum and value are based on deciles (the return difference between the top and bottom deciles) then the winner stock is getting  $10/N$  while the loser stock gets  $-10/N$ .
- The same applies to value versus growth stocks.
- Rule: the sum of the row corresponding to absolute views is one, while the sum of the row corresponding to relative views is zero.

# Computing the $\mu^v$ Vector

- It is the  $K \times 1$  vector of  $K$  views on expected returns.
- Using the absolute views above  $\mu^v = [0.05, 0.12]'$
- Using the relative views above, the first element is the payoff to momentum trading strategy (sample mean); the second element is the payoff to value investing (sample mean).

# The $\Omega$ Matrix

- $\Omega$  is a  $K \times K$  covariance matrix expressing uncertainty about views.
- It is typically assumed to be diagonal.
- In the absolute views case described above  $\Omega(1,1)$  denotes uncertainty about the first view while  $\Omega(2,2)$  denotes uncertainty about the second view – both are at the discretion of the econometrician/investor.

# The $\Omega$ Matrix

- In the relative views described above:  $\Omega(1,1)$  denotes uncertainty about momentum. This could be the sample variance of the momentum payoff.
- $\Omega(2,2)$  denotes uncertainty about the value payoff. This could be the sample variance of the value payoff.



# Deciding Upon $\tau$

- There are many debates among professionals about the right value of  $\tau$ .
- From a conceptual perspective it could be  $1/T$  where  $T$  denotes the sample size.
- You can pick  $\tau = 0.01$
- You can also use other values and examine how they perform in real-time investment decisions.



# Maximizing Sharpe Ratio

- The remaining task is to run the maximization program

$$\max_w \frac{w' \mu_{BL}}{\sqrt{w' \Sigma w}}$$

- Preferably, impose portfolio constraints, such that each of the  $w$  elements is bounded below by 0 and subject to some agreed upon upper bound, as well as the sum of the  $w$  elements is equal to one.

# Extending the BL Model to Incorporate Sample Moments

- Consider a sample of size  $T$ , e.g.,  $T = 60$  monthly observations.
- Let us estimate the mean and the covariance ( $V$ ) of our  $N$  assets based on the sample.
- Then the vector of expected return that serves as an input for asset allocation is given by

$$\mu = \left[ \Delta^{-1} + (V_{sample}/T)^{-1} \right]^{-1} \left[ \Delta^{-1} \mu_{BL} + (V_{sample}/T)^{-1} \mu_{sample} \right]$$

where

$$\Delta = [(\tau\Sigma)^{-1} + P'\Omega^{-1}P]^{-1}$$

# Session #10: Risk Management: Down Side Risk Measures

# Downside Risk

- **Downside risk** is the financial risk associated with losses.
- Downside risk measures quantify the risk of losses, whereas volatility measures are both about the upside and downside outcomes.
- Volatility treats symmetrically up and down moves (relative to the mean).
- Or volatility is about the entire distribution while down side risk concentrates on the left tail.

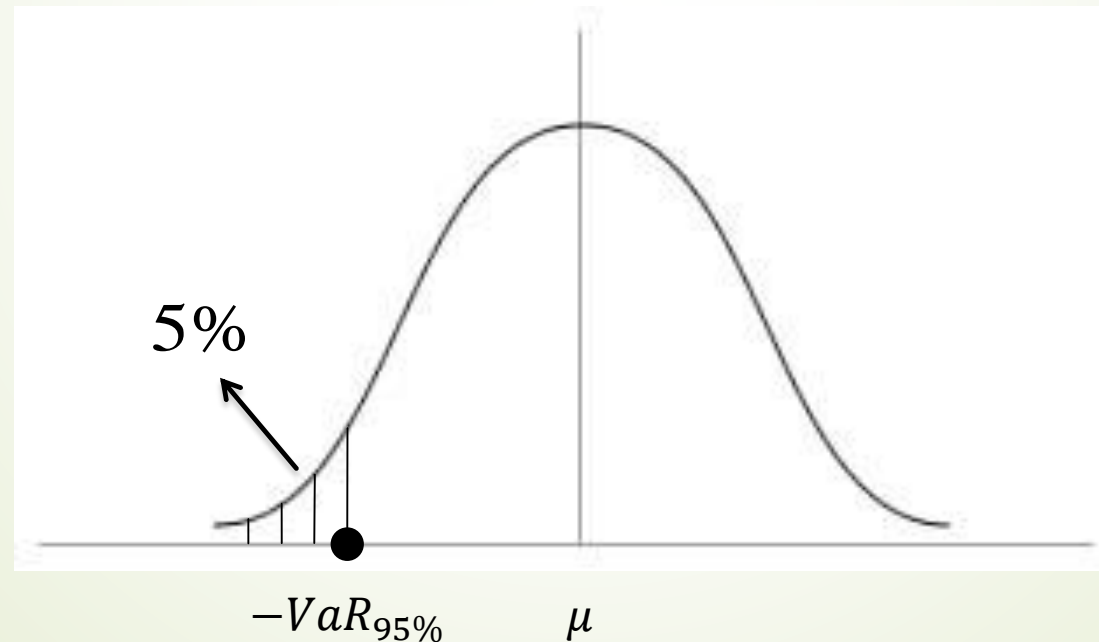
# Downside Risk

- Example of downside risk measures
  - Value at Risk (VaR)
  - Expected Shortfall
  - Semi-variance
  - Maximum drawdown
  - Downside Beta
  - Shortfall probability
  
- We will discuss below all these measures.

# Value at Risk (VaR)

- The  $VaR_{95\%}$  says that there is a 5% chance that the realized return, denoted by  $R$ , will be less than  $-VaR_{95\%}$ .
- More generally,

$$\Pr(R \leq -VaR_{1-\alpha}) = \alpha$$



# Value at Risk (VaR)

$$\frac{-VaR_{1-\alpha} - \mu}{\sigma} = \Phi^{-1}(\alpha) \Rightarrow VaR_{1-\alpha} = -(\mu + \sigma\Phi^{-1}(\alpha))$$

where

$\Phi^{-1}(\alpha)$ , the critical value, is the inverse cumulative distribution function of the standard normal evaluated at  $\alpha$ .

► Let  $\alpha = 5\%$  and assume that

$$R \sim N(\mu, \sigma^2)$$

► The critical value is  $\Phi^{-1}(0.05) = -1.64$



# Value at Risk (VaR)

Therefore

$$VaR_{1-\alpha} = -(\mu + \Phi^{-1}(\alpha)\sigma) = -\mu + 1.64\sigma$$

Check:

If

$$R \sim N(\mu, \sigma^2)$$

Then

$$\begin{aligned} \Pr(R \leq -VaR_{1-\alpha}) &= \Pr\left(\frac{R - \mu}{\sigma} \leq \frac{-VaR_{1-\alpha} - \mu}{\sigma}\right) \\ &= \Pr\left(\frac{R - \mu}{\sigma} < \frac{\mu - 1.64\sigma - \mu}{\sigma}\right) \\ &= \Pr(z < -1.64) = \Phi(-1.64) = 0.05 \end{aligned}$$

# Example: The US Equity Premium

➤ Suppose:

$$R \sim N(0.08, 0.20^2)$$

$$\Rightarrow VaR_{0.95} = -(0.08 - 1.64 \cdot 0.20) = 0.25$$

➤ That is to say that we are 95% sure that the future equity premium won't decline more than 25%.

➤ If we would like to be 97.5% sure – the price is that the threshold loss is higher:

$$VaR_{0.975} = -(0.08 - 1.96 \cdot 0.20) = 0.31$$

# VaR of a Portfolio

- Evidently, the VaR of a portfolio is not necessarily lower than the combination of individual VaR-s – which is apparently at odds with the notion of diversification.
- However, VaR is a downside risk measure while volatility (a risk measure) does diminish with better diversification.

# Backtesting the VaR

- The VaR requires the specification of the exact distribution and its parameters (e.g., mean and variance).
- Typically the normal distribution is chosen.
- Mean could be the sample average.
- Volatility estimates could follow ARCH, GARCH, EGARCH, stochastic volatility, and realized volatility, all of which are described later in this course.
- We can examine the validity of VaR using backtesting.

# Backtesting the VaR

- Assume that stock returns are normally distributed with mean and variance that vary over time

$$r_t \sim N(\mu_t, \sigma_t^2) \quad \forall t = 1, 2, \dots, T$$

- The sample is of length  $T$ .
- Receipt for backtesting is as follows.
- Use the first, say, sixty monthly observations to estimate the mean and volatility and compute the VaR.
- If the return in month 61 is below the VaR set an indicator function  $I$  to be equal to one; otherwise, it is zero.

# Backtesting the VaR

- Repeat this process using either a rolling or recursive schemes and compute the fraction of time when the next period return is below the VaR.
- If  $\alpha = 5\%$  - only 5% of the returns should be below the computed VaR.
- Suppose we get 5.5% of the time – is it a bad model or just a bad luck?



# Model Verification Based on Failure Rates

- To answer that question – let us discuss another example which requires a similar test statistic.
- Suppose that  $Y$  analysts are making predictions about the market direction for the upcoming year. The analysts forecast whether market is going to be up or down.
- After the year passes you count the number of wrong analysts. An analyst is wrong if he/she predicts up move when the market is down, or predict down move when the market is up.



# Model Verification Based on Failure Rates

- Suppose that the number of wrong analysts is  $x$ .
- Thus, the fraction of wrong analysts is  $P = x/Y$  – this is the failure rate.

# The Test Statistic

➤ The hypothesis to be tested is

$$H_0: P = P_0$$

*H<sub>1</sub>: Otherwise*

➤ Under the null hypothesis it follows that

$$f(x) = \binom{y}{x} P_0^x (1 - P_0)^{y-x}$$

# The Test Statistic

► Notice that

$$E(x) = P_0 y$$

$$VaR(x) = P_0(1 - P_0)y$$

Thus

$$Z = \frac{x - P_0 y}{\sqrt{P_0(1 - P_0)y}} \sim N(0,1)$$

# Back to Backtesting VaR: A Real Life Example

- In its 1998 annual report, JP Morgan explains: In 1998, daily revenue fell short of the downside (95% VaR) band on 20 trading days (out of 252) or more than 5% of the time ( $252 \times 5\% = 12.6$ ).
- Is the difference just a bad luck or something more systematic? We can test the hypothesis that it is a bad luck.

$$\begin{array}{l} H_0: x = 12.6 \\ H_1: \textit{Otherwise} \end{array} \rightarrow Z = \frac{20 - 12.6}{\sqrt{0.05 \cdot 0.95 \cdot 252}} = 2.14$$

# Back to Backtesting VaR: A Real Life Example

- Notice that you reject the null since 2.14 is higher than the critical value of 1.96.
- That suggests that JPM should search for a better model.
- They did find out that the problem was that the actual revenue departed from the normal distribution.

# Expected Shortfall (ES): Truncated Distribution

- ES is the expected value of the loss conditional upon the event that the actual return is below the VaR.
- The ES is formulated as

$$ES_{1-\alpha} = -E[R | R \leq -VaR_{1-\alpha}]$$

# Expected Shortfall (ES) and the Truncated Normal Distribution

- Assume that returns are normally distributed:

$$R \sim N(\mu, \sigma^2)$$

$$\Rightarrow ES_{1-\alpha} = -E[R | R \leq \mu + \Phi^{-1}(\alpha)\sigma]$$

$$\Rightarrow ES_{1-\alpha} = -\mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}$$



# Expected Shortfall (ES) and the Truncated Normal Distribution

where  $\phi(\cdot)$  is the pdf of the standard normal density

e.g.  $\phi(-1.64) = 0.103961$

This formula for ES is about the expected value of a truncated normally distributed random variable.

# Expected Shortfall (ES) and the Truncated Normal Distribution

Proof:

$$x \sim N(\mu, \sigma^2)$$

$$E(x|x \leq -VaR_{1-\alpha}) = \mu - \frac{\sigma\varphi(\kappa_0)}{\Phi(\kappa_0)} = \mu - \sigma \frac{\varphi(\Phi^{-1}(\alpha))}{\alpha}$$

since

$$\kappa_0 = \frac{-VaR_{1-\alpha} - \mu}{\sigma} = \Phi^{-1}(\alpha)$$

# Expected Shortfall: Example

Example:

$$\mu = 8\%$$

$$\sigma = 20\%$$

$$ES_{95\%} = -0.08 + 0.20 \frac{\varphi(-1.64)}{0.05} \approx -0.08 + 0.20 \frac{0.10}{0.05} = 0.32$$

$$ES_{97.5\%} = -0.08 + 0.20 \frac{\varphi(-1.96)}{0.025} \approx -0.08 + 0.20 \frac{0.06}{0.25} = 0.40$$

# Expected Shortfall (ES) and the Truncated Normal Distribution

- Previously, we got that the VaRs corresponding to those parameters are 25% and 31%.
- Now the expected losses are higher, 32% and 40%.
- Why?
- The first lower figures (VaR) are unconditional in nature relying on the entire distribution.
- In contrast, the higher ES figures are conditional on the existence of shortfall – realized return is below the VaR.

# Expected Shortfall in Decision Making

- The mean variance paradigm minimizes portfolio volatility subject to an expected return target.
- Suppose you attempt to minimize ES instead subject to expected return target.

# Expected Shortfall with Normal Returns

- If stock returns are normally distributed then the ES chosen portfolio would be identical to that based on the mean variance paradigm.
- No need to go through optimization to prove that assertion.
- Just look at the expression for ES under normality to quickly realize that you need to minimize the volatility of the portfolio subject to an expected return target.

# Target Semi Variance

- Variance treats equally downside risk and upside potential.
- The semi-variance, just like the VaR, looks at the downside.
- The target semi-variance is defined as:

$$\lambda(h) = E[\min(r - h, 0)^2]$$

where  $h$  is some target level.

- For instance,  $h = R_f$
- Unlike the variance,

$$\sigma^2 = E(r - \mu)^2$$



# Target Semi Variance

► The target semi-variance:

1. Picks a target level as a reference point instead of the mean.
2. Gives weight only to negative deviations from a reference point.

# Target Semi Variance

► Notice that if  $r \sim N(\mu, \sigma^2)$

$$\lambda(h) = \sigma^2 \frac{h - \mu}{\sigma} \varphi\left(\frac{h - \mu}{\sigma}\right) + \sigma^2 \left[ \left(\frac{h - \mu}{\sigma}\right)^2 + 1 \right] \Phi\left(\frac{h - \mu}{\sigma}\right)$$

where

$\phi$  and  $\Phi$  are the *PDF* and *CDF* of a  $N(0,1)$  variable, respectively

Of course if  $h = \mu$

$$\text{then } \lambda(h) = \frac{\sigma^2}{2}$$

# Maximum Drawdown (MD)

- The MD ( $M$ ) over a given investment horizon is the largest  $M$ -month loss of all possible  $M$ -month continuous periods over the entire horizon.
- Useful for an investor who does not know the entry/exit point and is concerned about the worst outcome.
- It helps determine the investment risk.

# Down Size Beta

- I will introduce three distinct measures of downsize beta – each of which is valid and captures the down side of investment payoffs.
- Displayed are the population betas.
- Taking the formulations into the sample – simply replace the expected value by the sample mean.

# Downside Beta

$$\beta_{im}^{(1)} = \frac{E \left[ (R_i - R_f) \left[ \min \left( (R_m - R_f), 0 \right) \right] \right]}{E \left[ \min \left( (R_m - R_f), 0 \right) \right]^2}$$

- The numerator in the equation is referred to as the co-semi-variance of returns and is the covariance of returns below  $R_f$  on the market portfolio with return in excess of  $R_f$  on security  $i$ .
- It is argued that risk is often perceived as downside deviations below a target level by market participants and the risk-free rate is a replacement for average equity market returns.

# Downside Beta

$$\beta_{im}^{(2)} = \frac{E[(R_i - \mu_i) \min(R_m - \mu_m, 0)]}{E[\min(R_m - \mu_m, 0)]^2}$$

where  $\mu_i$  and  $\mu_m$  are security  $i$  and market average return respectively.

One can modify the down side beta as follows:

$$\beta_{im}^{(3)} = \frac{E[\min(R_i - \mu_i, 0) \min(R_m - \mu_m, 0)]}{E[\min(R_m - \mu_m, 0)]^2}$$

# Shortfall Probability

- We now turn to understand the notion of shortfall probability.
- While VaR specifies upfront the probability of undesired outcome and then finds the threshold level, shortfall probability gives a threshold level and seeks for the probability that the outcome is below that threshold.
- We will thoroughly study the implications of shortfall probability for long horizon investment decisions.



# Shortfall Probability

## in Long Horizon Asset Management

- ▶ Let us denote by  $R$  the cumulative return on the investment over several years (say  $T$  years).
- ▶ Rather than finding the distribution of  $R$  we analyze the distribution of

$$r = \ln(1 + R)$$

which is the continuously compounded return over the investment horizon.

# Shortfall Probability

## in Long Horizon Asset Management

- The investment value after  $T$  years is

$$V_T = V_0(1 + R_1)(1 + R_2) \dots (1 + R_T)$$

- Dividing both sides of the equation by  $V_0$  we get

$$\frac{V_T}{V_0} = (1 + R_1)(1 + R_2) \dots (1 + R_T)$$

- Thus

$$1 + R = (1 + R_1)(1 + R_2) \dots (1 + R_T)$$

# Shortfall Probability

## in Long Horizon Asset Management

- ▶ Taking natural log from both sides we get

$$r = r_1 + r_2 + \cdots + r_T$$

- ▶ Assuming that

$$r_t \stackrel{IID}{\sim} N(\mu, \sigma^2) \quad \forall t = 1, \dots, T$$

- ▶ Then using properties of the normal distribution, we get

$$r \sim N(T\mu, T\sigma^2)$$

# Shortfall Probability and Long Horizon

- The normality assumption for log return implies the log normal distribution for the cumulative return – more later.
- Let us understand the concept of shortfall probability.
- We ask: what is the probability that the investment yields a return smaller than a threshold level (e.g., the risk-free rate)?
- To answer this question we need to compute the value of a risk-free investment over the  $T$  year period.

# Shortfall Probability and Long Horizon

► The value of such a risk-free investment is

$$\begin{aligned}V_{r_f} &= V_0(1 + R_f)^T \\ &= V_0 \exp(T r_f)\end{aligned}$$

where  $r_f$  is the continuously compounded risk free rate.

# Shortfall Probability and Long Horizon

- Essentially we ask: what is the probability that

$$V_T < V_{rf}$$

- This is equivalent to asking what is the probability that

$$\frac{V_T}{V_0} < \frac{V_{rf}}{V_0}$$

- This, in turn, is equivalent to asking what is the probability that

$$\ln \left( \frac{V_T}{V_0} \right) < \ln \left( \frac{V_{rf}}{V_0} \right)$$

# Shortfall Probability and Long Horizon

► So we need to work out

$$p(r < Tr_f)$$

► Subtracting  $T\mu$  and dividing by  $\sqrt{T}\sigma$  both sides of the inequality we get

$$P\left(z < \sqrt{T} \left(\frac{r_f - \mu}{\sigma}\right)\right)$$

► We can denote this probability by

$$\text{Shortfall probability} = \Phi\left(\sqrt{T} \left(\frac{r_f - \mu}{\sigma}\right)\right)$$



# Shortfall Probability and Long Horizon

- Typically  $r_f < \mu$  which means the probability diminishes the larger  $T$ .
- Notice that the shortfall probability can be written as a function of the Sharpe ratio of log returns:

$$SP = \Phi(-\sqrt{T} SR)$$

# Example

- Take  $r_f=0.04$ ,  $\mu=0.08$ , and  $\sigma=0.2$  per year. What is the Shortfall Probability for investment horizons of 1, 2, 5, 10, and 20 years?
- Use the excel norm.dist function.
- If  $T=1$  SP=0.42
- If  $T=2$  SP=0.39
- If  $T=5$  SP=0.33
- If  $T=10$  SP=0.26
- If  $T=20$  SP=0.19

# Cost of Insuring against Shortfall

- ▶ Let us now understand the mathematics of insuring against shortfall.
- ▶ Without loss of generality let us assume that
$$V_0 = 1$$
- ▶ The investment value at time  $T$  is given by the random variable  $V_T$

# Cost of Insuring against Shortfall

- Once we insure against shortfall the investment value after  $T$  years becomes
  - If  $V_T > \exp(T r_f)$  you get  $V_T$
  - If  $V_T < \exp(T r_f)$  you get  $\exp(T r_f)$

# Cost of Insuring against Shortfall

- So you essentially buy an insurance policy
- The policy pays 0 if  $V_T > \exp(T r_f)$  while it pays  $\exp(T r_f) - V_T$  if  $V_T < \exp(T r_f)$
- You ultimately need to price a contract with terminal payoff given by

$$\max \{0, \exp(T r_f) - V_T\}$$

# Cost of Insuring against Shortfall

► This is a European put option expiring in  $T$  years with

1.  $S=1$
2.  $K = \exp(T r_f)$ .
3. Risk-free rate given by  $r_f$
4. Volatility given by  $\sigma$
5. Dividend yield given by  $\delta = 0$

# Cost of Insuring against Shortfall

➤ The B&S formula indicates that

$$Put = K \exp(-Tr_f) N(-d_2) - S \exp(-\delta T) N(-d_1)$$

➤ Given the parameter outlined above the put price becomes

$$Put = N\left(\frac{1}{2}\sigma\sqrt{T}\right) - N\left(-\frac{1}{2}\sigma\sqrt{T}\right)$$



# Cost of Insuring against Shortfall

To show the pricing formula of the put use the following:

$$d_1 = \frac{\ln(S/K) + (r - \delta + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T}$$

while  $N(-d_1) = 1 - N(d_1)$

$$N(-d_2) = 1 - N(d_2)$$

# Cost of Insuring against Shortfall

➔ The B&S option-pricing model gives the current put price  $P$  as

$$Put = N(d_1) - N(d_2)$$

where

$$d_1 = \frac{\sigma\sqrt{T}}{2}$$

$$d_2 = -d_1$$

and  $N(d)$  is  $prob\{z < d\}$

# Cost of Insuring against Shortfall

➔ For  $\sigma = 0.2$  (per year)

<b>T (years)</b>	<b>P</b>
1	0.08
5	0.18
10	0.25
20	0.35
30	0.42
50	0.52

# Cost of Insuring against Shortfall

- We have found out that the cost of the insurance increases in  $T$ , even when the probability of shortfall decreases in  $T$  (as long as the Sharpe ratio is positive).
- To get some idea about this apparently surprising outcome it would be essential to discuss the expected value of the investment payoff given the shortfall event.
- It is a great opportunity to understand down side risk when the underlying distribution is log normal rather than normal.

# The Expected Value of Cumulative Return

- Two proper questions emerge at this stage:
  1. What is the expected value of cumulative return during the investment horizon  $E(V_T)$ ?
  2. What is the conditional expectation – conditional on shortfall  $E[V_T \mid V_T < \exp(Tr_f)]$  ?
- We assume, without loss of generality, that the initial invested wealth is one.

# The Expected Value of Cumulative Return

- Notice that, given the tools we have acquired thus far, finding the conditional expectation is a nontrivial task since  $V_T$  is not normally distributed – rather it is log-normally distributed since.

$$\ln(V_T) \sim N(T\mu, T\sigma^2)$$

- Thus, let us first display some properties of the log normal distribution.

# The Log Normal Distribution

- Suppose that  $x$  has the log normal distribution. Then the parameters  $\mu$  and  $\sigma$  are, respectively, the mean and the standard deviation of the variable's natural logarithm, which means

$$x = e^{\mu + \sigma z}$$

where  $z$  is a standard normal variable.

- The probability density function of a log-normal distribution is



# The Log Normal Distribution

$$f_x(x, \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, x > 0$$

- If  $x$  is a log-normally distributed variable, its expected value and variance are given by

$$E[x] = e^{\mu + \frac{1}{2}\sigma^2}$$

$$\text{Var}[x] = (e^{\sigma^2} - 1) e^{2\mu + \sigma^2} = (e^{\sigma^2} - 1) (E[x])^2$$

# The Mean and Variance of $V_T$

- Using moments of the log normal distribution, the mean and variance of  $V_T$  are

$$E(V_T) = \exp\left(T\mu + \frac{1}{2}T\sigma^2\right)$$

$$\text{Var}(V_T) = \exp(2T\mu + T\sigma^2)(\exp(T\sigma^2) - 1)$$

- Next, we aim to find the conditional mean.

# The Mean of a Variable that has the Truncated Log Normal Distribution

$$\bar{F}(c)E(x|x > c) = \int_c^{\infty} x \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\ln(x)-\mu}{\sigma}\right]^2} dx$$

where  $1/\bar{F}(c)$  is a normalizing constant.

► Let us make change of variables:

$$\frac{\ln(x) - \mu}{\sigma} = t \Rightarrow x = e^{t\sigma + \mu} \text{ and } dx = \sigma e^{t\sigma + \mu} dt$$

# The Mean of a Variable that has the Truncated Log Normal Distribution

then:

$$\begin{aligned}\bar{F}(c)E(x|x > c) &= \int_{\frac{\ln(c)-\mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\Pi}} e^{-\frac{1}{2}t^2} \sigma e^{t\sigma+\mu} dt = \\ &= \frac{1}{\sqrt{2\Pi}} \int_{\frac{\ln(c)-\mu}{\sigma}}^{\infty} e^{-\frac{1}{2}t^2+t\sigma+\mu} dt \\ &= \frac{1}{\sqrt{2\Pi}} \int_{\frac{\ln(c)-\mu}{\sigma}}^{\infty} e^{-\frac{1}{2}(t-\sigma)^2+(\mu+0.5\sigma^2)} dt \\ &= e^{(\mu+0.5\sigma^2)} \frac{1}{\sqrt{2\Pi}} \int_{\frac{\ln(c)-\mu}{\sigma}}^{\infty} e^{-\frac{1}{2}(t-\sigma)^2} dt\end{aligned}$$

# The Mean of a Variable that has the Truncated Log Normal Distribution

- ▶ Let us make another change of variables:

$$v = t - \sigma \Rightarrow dv = dt$$

$$\bar{F}(c)E(x|x > c) = e^{(\mu+0.5\sigma^2)} \frac{1}{\sqrt{2\pi}} \int_{\frac{\ln(c)-\mu}{\sigma}-\sigma}^{\infty} e^{-\frac{1}{2}v^2} dv$$

- ▶ The integral is the complement *CDF* of the standard normal random variable.

# The Mean of a Variable that has the Truncated Log Normal Distribution

► Thus, the formula is reduced to:

$$\begin{aligned}\bar{F}(c)E_{LN}(x|x > c) &= e^{(\mu+0.5\sigma^2)} \left[ 1 - \Phi \left( \frac{\ln(c) - \mu - \sigma^2}{\sigma} \right) \right] \\ &= e^{(\mu+0.5\sigma^2)} \Phi \left( \frac{-\ln(c) + \mu + \sigma^2}{\sigma} \right)\end{aligned}$$

# The Mean of a Variable that has the Truncated Log Normal Distribution

► In the same way we can show that:

$$\begin{aligned}\bar{F}(c) \cdot E_{\text{LN}}(X|X \leq c) &= \int_0^c x \frac{1}{x\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\ln(x)-\mu}{\sigma}\right]^2} dx = \int_{-\infty}^{\frac{\ln(c)-\mu}{\sigma}} e^{-\frac{1}{2}t^2+t\sigma+\mu} dt \\ &= e^{(\mu+0.5\sigma^2)} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\ln(c)-\mu}{\sigma}} e^{-\frac{1}{2}(t-\sigma)^2} dt = e^{(\mu+0.5\sigma^2)} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\ln(c)-\mu}{\sigma}-\sigma} e^{-\frac{1}{2}v^2} dv \\ &= e^{(\mu+0.5\sigma^2)} \cdot \Phi\left(\frac{\ln(c)-\mu-\sigma^2}{\sigma}\right)\end{aligned}$$



# Punch Lines

$$E_{\text{LN}}(x|x > c) = E_{\text{LN}}(x) \cdot \frac{\Phi\left(\frac{-\ln(c) + \mu + \sigma^2}{\sigma}\right)}{\Phi\left(\frac{-\ln(c) + \mu}{\sigma}\right)}$$

$$E_{\text{LN}}(x|x \leq c) = E_{\text{LN}}(x) \cdot \frac{\Phi\left(\frac{\ln(c) - \mu - \sigma^2}{\sigma}\right)}{\Phi\left(\frac{\ln(c) - \mu}{\sigma}\right)}$$

# The Expected Value given Shortfall

► The expected value given shortfall is

$$E[V_T | \text{shortfall}] = e^{\left(T\mu + \frac{1}{2}T\sigma^2\right)} \cdot \frac{\Phi\left(\frac{Tr_f - T\mu - T\sigma^2}{\sqrt{T}\sigma}\right)}{\Phi\left(\frac{Tr_f - T\mu}{\sqrt{T}\sigma}\right)}$$

or

$$E[V_T | \text{shortfall}] = e^{\left(T\mu + \frac{1}{2}T\sigma^2\right)} \cdot \frac{\Phi\left(-\sqrt{T}(SR + \sigma)\right)}{\Phi\left(-\sqrt{T} \cdot SR\right)}$$

# The Expected Value given Shortfall

➤ Thus,

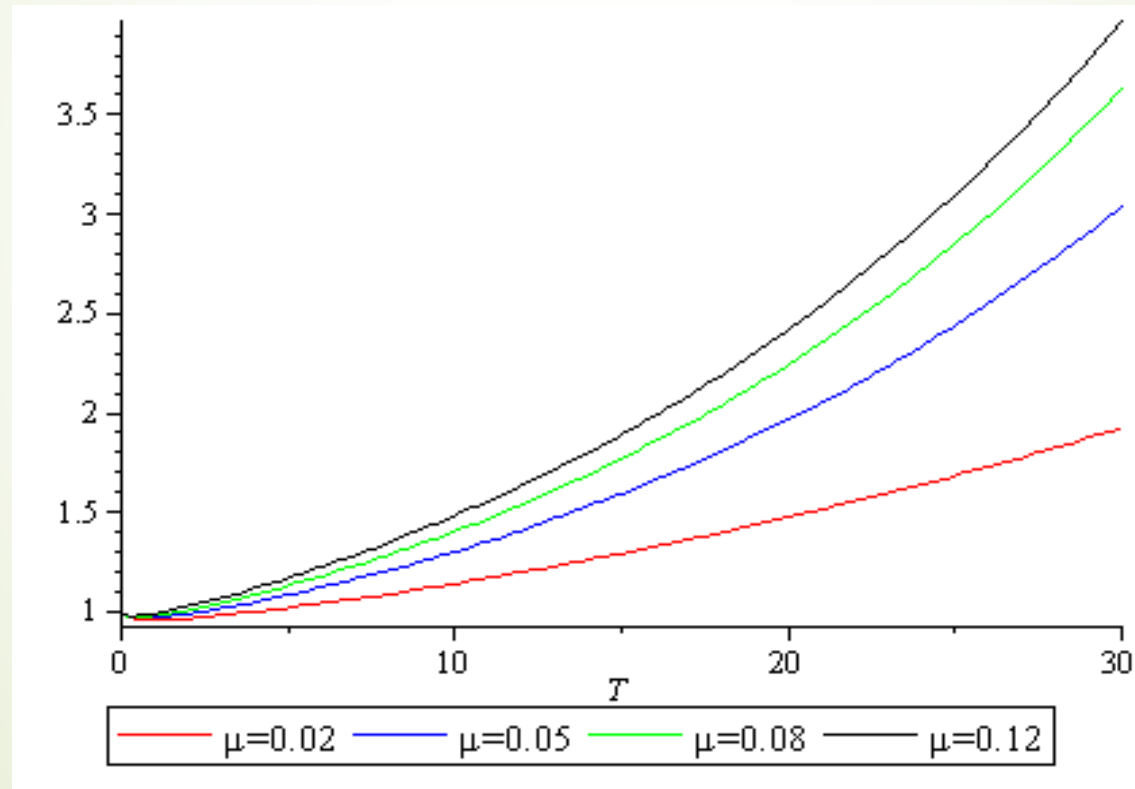
$$\text{Prob}(\text{shortfall}) \times E[V_T | \text{shortfall}] = E[V_t] \Phi\left(-\sqrt{T}(SR + \sigma)\right)$$

- Which means that the shortfall probability times the expected value given shortfall is equal to the unconditional expected value times a factor smaller than one.
- That factor diminishes with higher Sharpe ratio and/or with higher volatility.

# The Horizon Effect

## Numerical example

- ▶ Let's take  $r_f = 5\%$ ,  $\sigma = 10\%$ . For different values of  $\mu > r_f$  the conditional expectation over horizon  $T$  looks like:



# The Horizon Effect

- Previously we have shown that even when the shortfall probability diminishes with the investment horizon, the cost of insuring against shortfall rises.
- Notice that the insured amount is
$$\exp(T r_f) - V_T$$
- The expected value of that insured amount given shortfall sharply rises with the investment horizon, which explains the increasing value of the put option.

# Value at Risk with Log Normal Distribution

- We have analyzed VaR when quantities of interest are normally distributed.
- It is challenging to extend the analysis to the case wherein the log normal distribution is considered.
- Analytics follow.

# VaR with Log Normal Distribution

- We are looking for threshold, VaR, such that

$$\alpha = \Pr (V_T | V_0 < VaR | V_0) = CDF(VaR | V_0)$$

- Then in order to find the threshold we need to calculate quantile of lognormal distribution:

$$\begin{aligned} VaR &= V_0 \cdot CDF^{-1}(\alpha; T\mu, T\sigma^2) \\ &= V_0 \cdot e^{T\mu + \sqrt{T}\sigma \cdot \Phi^{-1}(\alpha)} \end{aligned}$$

where  $\Phi^{-1}(\alpha)$  is as defined earlier.



# VaR with Log Normal Distribution

➔ Specifically,

$$\alpha = \Pr (V_T | V_0 < VaR | V_0) = \Pr (\ln (V_T | V_0) < \ln (VaR | V_0))$$

$$= \Pr \left( \frac{\ln (V_T | V_0) - T\mu}{\sqrt{T}\sigma} < \frac{\ln (VaR | V_0) - T\mu}{\sqrt{T}\sigma} \right)$$

$$= \Phi \left( \frac{\ln (VaR | V_0) - T\mu}{\sqrt{T}\sigma} \right)$$

# VaR with Log Normal Distribution

and then

$$\left( \frac{\ln (VaR | V_0) - T\mu}{\sqrt{T}\sigma} \right) = \Phi^{-1}(\alpha)$$

$$VaR = V_0 e^{T\mu + \sqrt{T}\sigma \cdot \Phi^{-1}(\alpha)}$$

- That is to say that there is a  $\alpha\%$  probability that the investment value at time  $T$  will be below that VaR.

# VaR with $t$ Distribution

- Suppose now that stock returns have a  $t$  distribution with  $\nu$  degrees of freedom and expected return and volatility given by  $\mu$  and  $\sigma$
- The pdf of stock return is formulated as

$$f(x|\mu, \sigma, \nu) = \frac{1}{\sigma\sqrt{\nu} \cdot B(\nu/2, 1/2)} \cdot \left(1 + \frac{(x - \mu)^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

# Partial Expectation

- Let  $Y = \frac{x-\mu}{\sigma}$ -standardized r.v distribution with  $F(x|0,1, v)$ .

Then,

$$PE(X|X \leq z) = \int_{-\infty}^z x \cdot f(x, v) dx = \int_{-\infty}^z \frac{1}{\sqrt{v} \cdot B(v/2, 1/2)} \cdot \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}} dx =$$

$$= \frac{1}{\sqrt{v} \cdot B(v/2, 1/2)} \cdot \left[ \frac{-v}{v-1} \left(1 + \frac{x^2}{v}\right)^{-\frac{v-1}{2}} \right]_{-\infty}^z =$$

$$\frac{1}{\sqrt{v} \cdot B(v/2, 1/2)} \cdot \frac{-v}{v-1} \left(1 + \frac{z^2}{v}\right)^{-\frac{v+1}{2}+1} = -f(z, v) \cdot \frac{v + z^2}{v-1}$$

# Partial Expectation

And thus

$$ES_T(\alpha) = \mu - \sigma \frac{f(q_{1-\alpha}, v)}{q_{1-\alpha}} \cdot \frac{v + q_{1-\alpha}^2}{v - 1}$$

Where  $q_x = VaR(1 - x)$  is  $x$  quantile of  $T \sim t_n$

# Long Run Return

## when Periodic Return has the $t$ -Distribution

- ▶ The sum of independent  $t$ -distributed random variables is not  $t$ -distributed. So we have no nice formula for expected shortfall in the long run. However, it can be approximated by normal with zero mean variance:

$$r_{approx.} \sim N\left(\mu, \sqrt{\frac{v}{v-2}}\sigma\right) \text{ for } v \geq 2$$

# Long Run Return

## when Periodic Return has the $t$ -Distribution

- Approximation makes sense for large  $v$ 's when  $t$  coincides with normal distribution.
- However, simulation studies show that for sufficient number of periods this approximation works well enough.



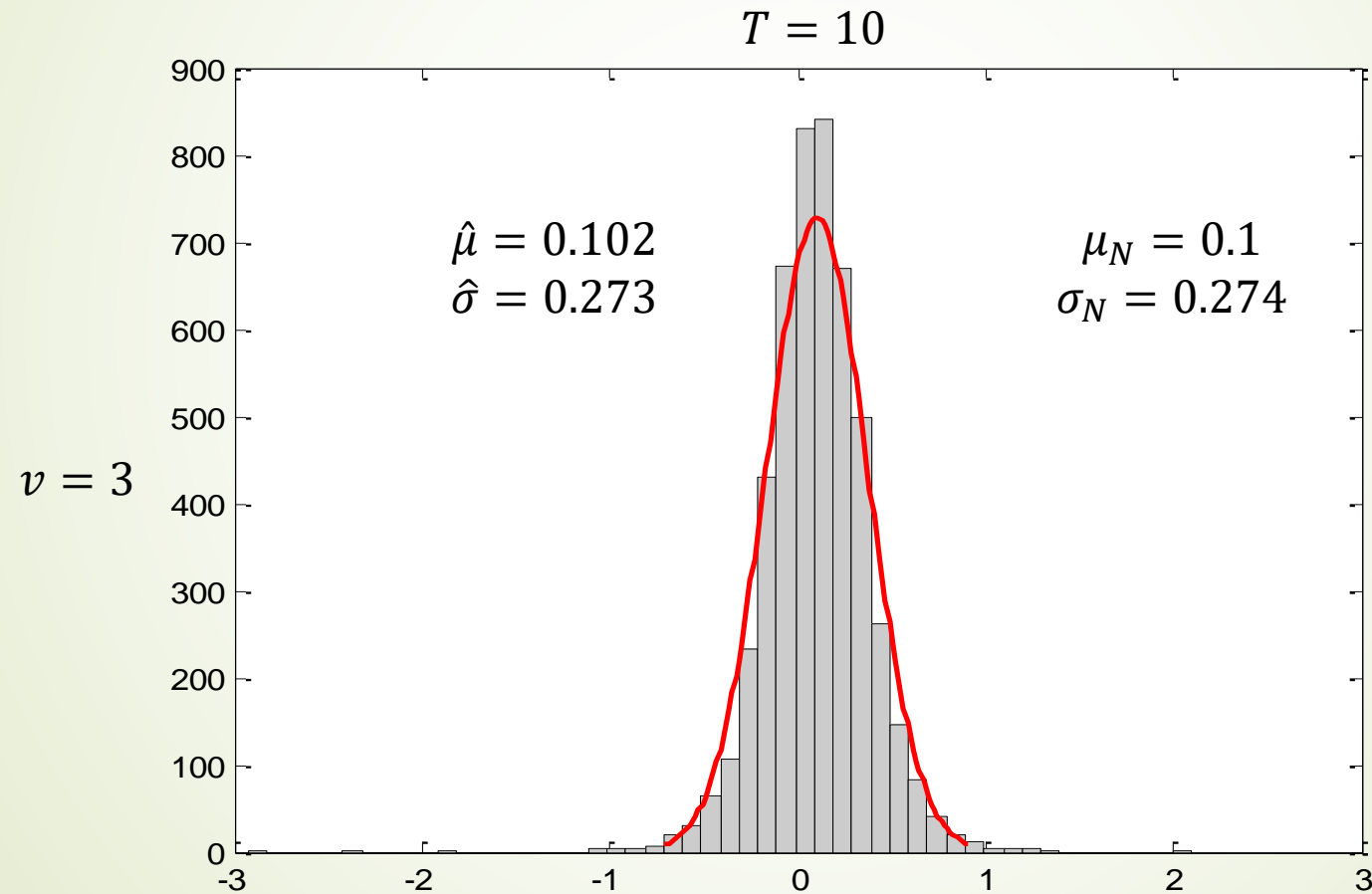
# Long Run Return

## when Periodic Return has the $t$ -Distribution

- ▶ However simulations shows that for sufficient number of periods this approximation works well enough.
- ▶ Let  $\mu_t = 0.01$ ;  $\sigma_t = 0.05$ . The next graphs show normal curve fit to the sum of  $t$  r.v.s (over  $T$  periods); sample estimates vs. predicted parameters are includes

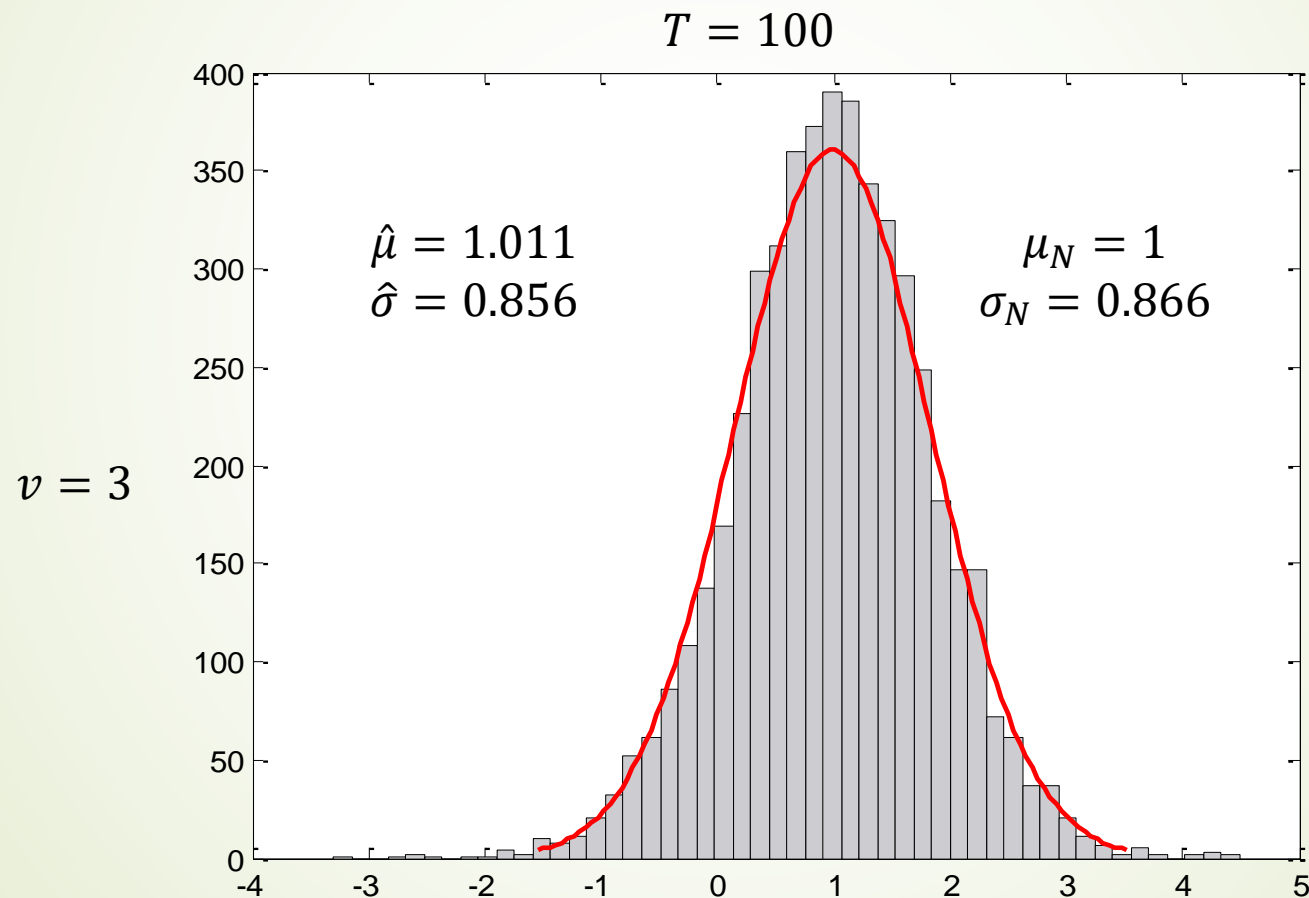
# Long Run Return

## when Periodic Return has the $t$ -Distribution

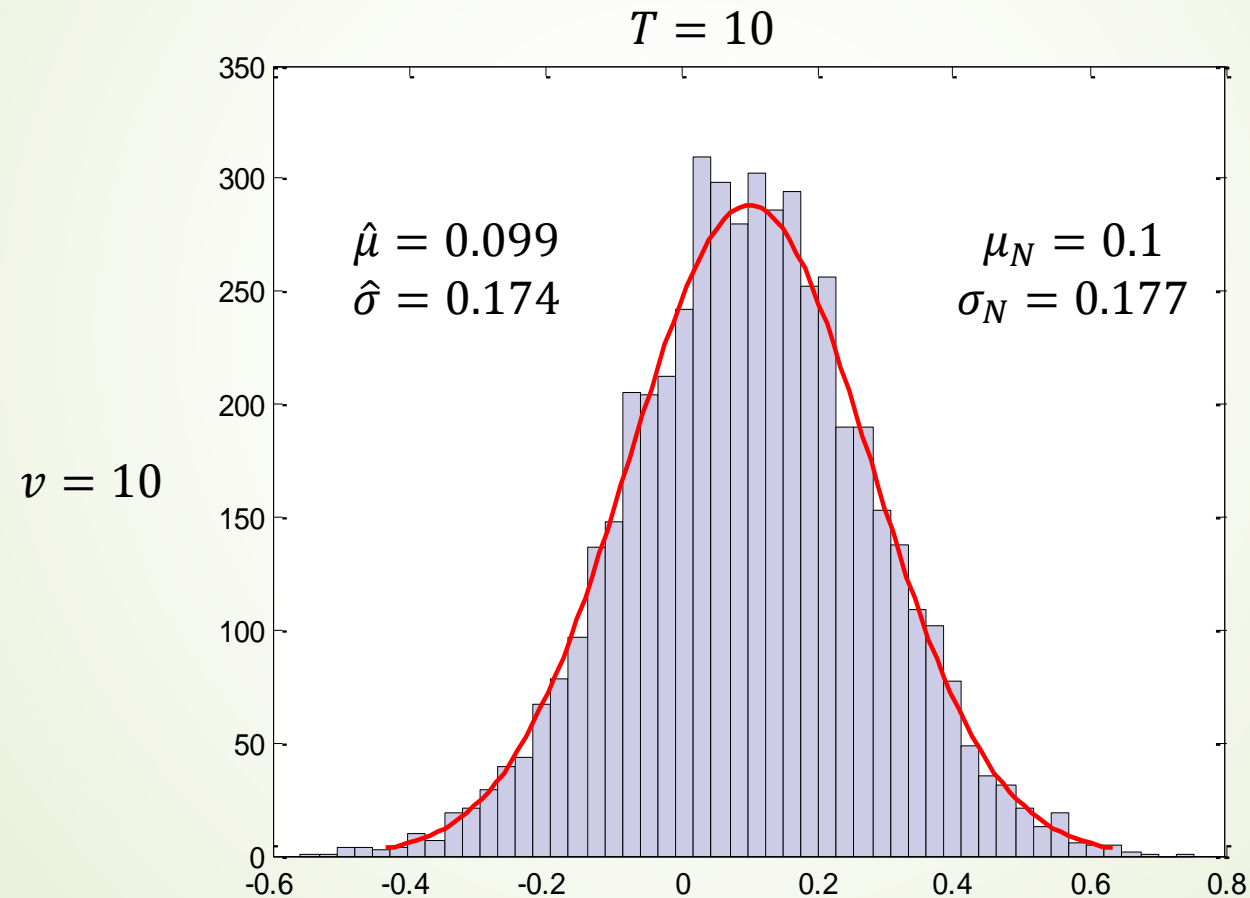


# Long Run Return

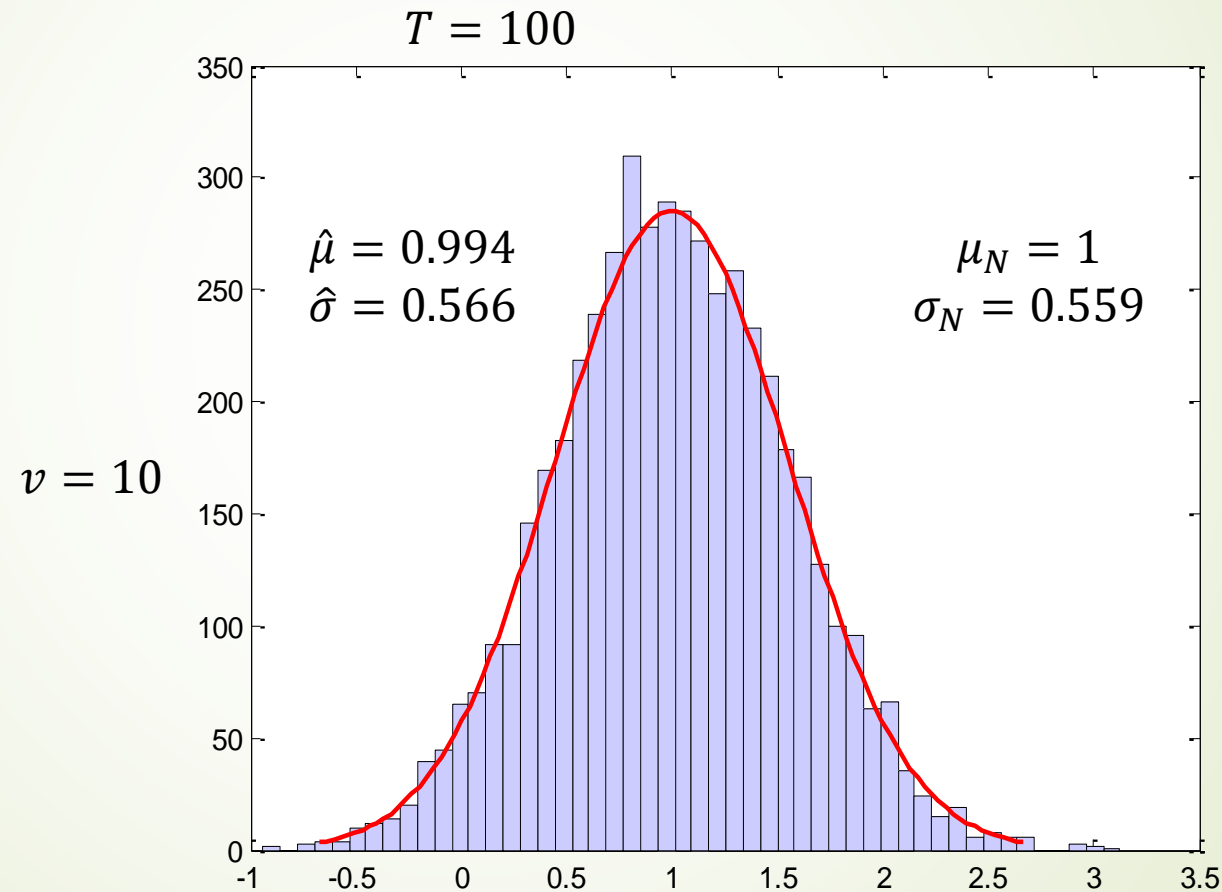
## when Periodic Return has the $t$ -Distribution



# Long Run Return when Periodic Return has the $t$ -Distribution



# Long Run Return when Periodic Return has the $t$ -Distribution



# Session #11 (part a): Testing the Black Scholes Formula

# Option Pricing

- The B&S call Option price is given by

$$C(S, K, \sigma, r, T, \delta) = Se^{-\delta T} N(d_1) - Ke^{-rT} N(d_2)$$

- The put Option price is

$$P(S, K, \sigma, r, T, \delta) = Ke^{-rT} N(-d_2) - Se^{-\delta T} N(-d_1)$$

$$\text{where } d_1 = \frac{\ln(S/K) + (r - \delta + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \text{ and } d_2 = d_1 - \sigma\sqrt{T}$$



# Option Pricing

➔ There are six inputs required:

$S$  - Current price of the underlying asset.

$K$  - Exercise/Strike price.

$r$  - Continuously compounded risk-free rate.

$T$  - Time to expiration.

$\sigma$  - Volatility.

$\delta$  - Continuously compounded dividend yield.

# The B&S Economy

- The B&S formula is derived under several assumptions:
  - The stock price follows a geometric Brownian motion (continuous path and continuous time).
  - The dividend is paid continuously and uniformly over time.
  - The interest rate is constant over time.

# The B&S Economy

- The underlying asset volatility is constant over time and it does not change with the option maturity or with the strike price.
- You can short sell or long any amount of the stock.
- You can borrow and lend in the risk-free rate.
- There are no transactions costs.

# Testing the B&S Formula

- Mark Rubinstein analyzes call options that are deep out of the money.
- He considers matched pairs: options with the same striking price, on the same underlying asset (stock), but with different time to maturity (expiration date).
- He examines overall 373 pairs.
- If B&S is correct then the implied volatility (IV) of the matched pair is equal. Time to maturity plays no role.

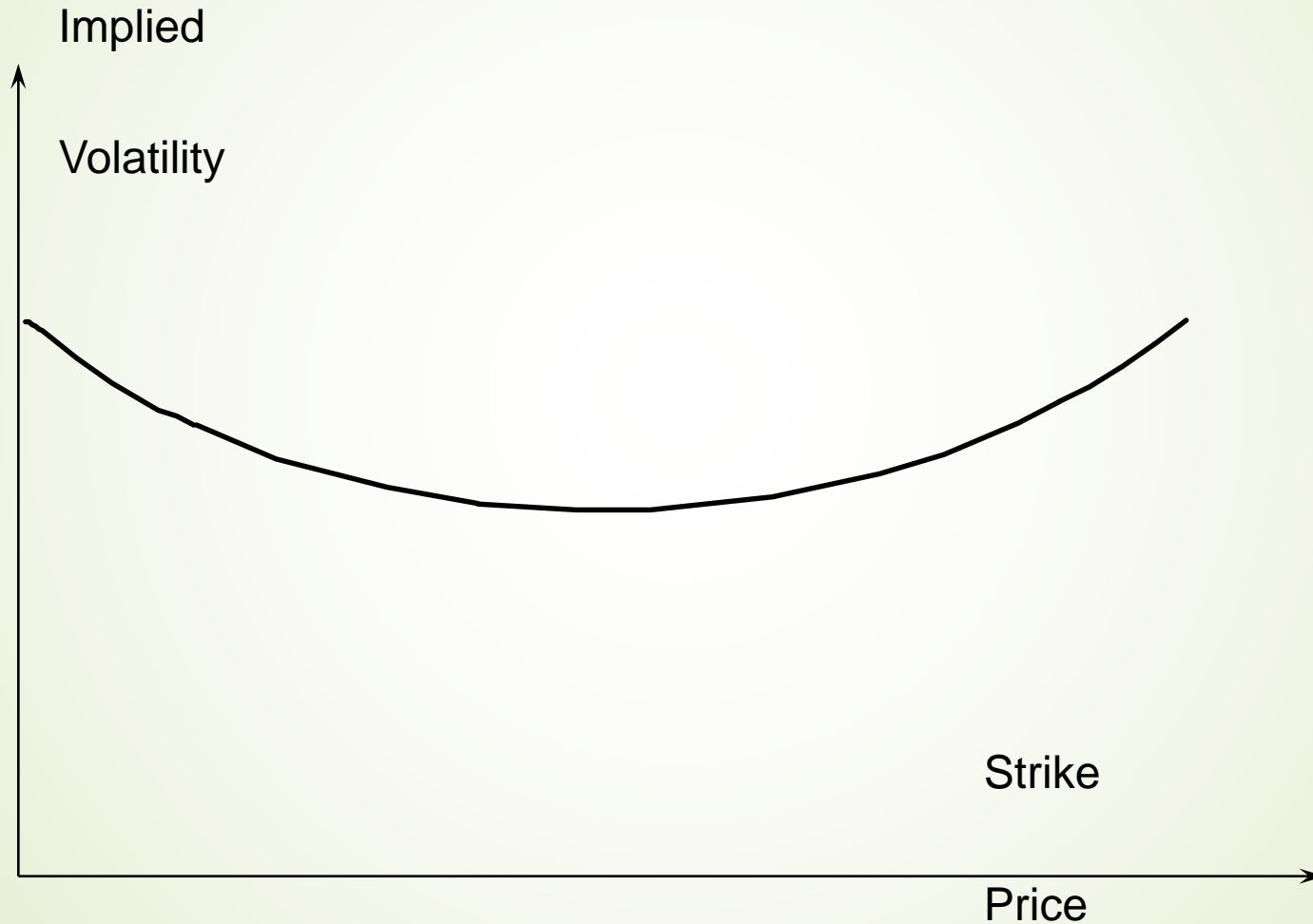
# Testing the B&S Formula

- However, Rubinstein finds that out of the 373 examined matched pairs – shorter maturity options had higher IV.
- Under the null – the expected value of such an outcome is  $373/2=186.5$ .
- Is the difference statistically significant?

# The Failure Rate based Test Statistic

- Use the failure rate test developed earlier to show that time to expiration does play a major role.
- That is to say that the constant volatility assumption is strongly violated in the data.

# The Volatility Smile for Foreign Currency Options





# Implied Distribution for Foreign Currency Options

- Both tails are heavier than the lognormal distribution.
- It is also “more peaked” than the lognormal distribution.

# The Volatility Smile for Equity Options



# Implied Distribution for Equity Options

- The left tail is heavier and the right tail is less heavy than the lognormal distribution.

# Ways of Characterizing the Volatility Smiles

- Plot implied volatility against  $K/S_0$  (The volatility smile is then more stable).
- Plot implied volatility against  $K/F_0$  (Traders usually define an option as at-the-money when  $K$  equals the forward price,  $F_0$ , not when it equals the spot price  $S_0$ ).
- Plot implied volatility against delta of the option (This approach allows the volatility smile to be applied to some non-standard options. At-the money is defined as a call with a delta of 0.5 or a put with a delta of  $-0.5$ . These are referred to as 50-delta options).

# Possible Causes of Volatility Smile

- Asset price exhibits jumps rather than continuous changes.
- Volatility for asset price is stochastic:
  - In the case of an exchange rate volatility is not heavily correlated with the exchange rate. The effect of a stochastic volatility is to create a symmetrical smile.
  - In the case of equities volatility is negatively related to stock prices because of the impact of leverage. This is consistent with the skew that is observed in practice.

# Volatility Term Structure

- In addition to calculating a volatility smile, traders also calculate a volatility term structure.
- This shows the variation of implied volatility with the time to maturity of the option.
- The volatility term structure tends to be downward sloping when volatility is high and upward sloping when it is low

# Example of a Volatility Surface

$$K/S_0$$

	<b>0.90</b>	<b>0.95</b>	<b>1.00</b>	<b>1.05</b>	<b>1.10</b>
1 month	14.2	13.0	12.0	13.1	14.5
3 months	14.0	13.0	12.0	13.1	14.2
6 months	14.1	13.3	12.5	13.4	14.2
1 year	14.7	14.0	13.5	14.0	14.8
2 years	15.0	14.4	14.0	14.5	15.1
5 years	14.8	14.6	14.4	14.7	15.0



# Session # 1 1 (part b): Time Varying Volatility Models

# Volatility Models

- We describe several volatility models commonly applied in analyzing quantities of interest in finance and economics:
  - ARCH
  - GARCH
  - EGARCH
  - Stochastic Volatility
  - Realized and implied Volatility

# Volatility Models

- All such models attempt to capture the empirical evidence that volatility is time varying (rather than constant) as well as persistent.
- The EGARCH captures the asymmetric response of volatility to advancing versus diminishing markets.
- In particular, volatility tends to be higher (lower) during down (up) markets.

# ARCH(1)

$$r_t = \mu + \varepsilon_t$$

$$\varepsilon_t = \sigma_t e_t \text{ where } e_t \sim N(0,1)$$

$$\sigma_t^2 = w + \alpha \varepsilon_{t-1}^2$$

$$E_{t-1}(\varepsilon_t^2) = E_{t-1}(\sigma_t^2 e_t^2) = \sigma_t^2$$

so

$\sigma_t^2$  is the conditional variance.

# ARCH(1)

$E(\varepsilon_{t-1}^2) = \bar{\sigma}^2$  is the unconditional variance.

$$\begin{aligned} E(\sigma_t^2) &= E(w + \alpha\varepsilon_{t-1}^2) \\ &= w + \alpha E(\varepsilon_{t-1}^2) \\ &= w + \alpha E(\sigma_{t-1}^2) E(e_{t-1}^2) \\ &= w + \alpha E(\sigma_{t-1}^2) \end{aligned}$$

$$\Rightarrow \bar{\sigma}^2 = E(\sigma_t^2) = E(\sigma_{t-1}^2) = E(\sigma_{t-2}^2)$$

$$E(\sigma_t^2) = \frac{w}{1-\alpha}$$

# ARCH(1)

► Fat tail?

$$\begin{aligned}\mu_4 &= \frac{E(\varepsilon_t^4)}{E(\varepsilon_t^2)^2} = \frac{E[E_{t-1}(\varepsilon_t^4)]}{E[E_{t-1}(e_t^2 \sigma_t^2)]^2} \\ &= \frac{E[E_{t-1}(e_t^4)(\sigma_t^4)]}{E[E_{t-1}(e_t^2 \sigma_t^2)]^2} = \frac{E(3\sigma_t^4)}{(E(\sigma_t^2))^2} \\ &= 3 \frac{E(\sigma_t^4)}{(E(\sigma_t^2))^2} \geq 3\end{aligned}$$

# ARCH(1)

► The last step follows because:

$$\text{Var}(\varepsilon_t^2) = E(\varepsilon_t^4) - E(\varepsilon_t^2)^2 \geq 0$$

SO

$$\frac{E(\varepsilon_t^4)}{E(\varepsilon_t^2)^2} \geq 1$$

► Yes – fat tail!



# GARCH(1,1)

$$r_t = \mu + \varepsilon_t$$

$$\varepsilon_t = \sigma_t e_t \text{ where } e_t \sim N(0,1)$$

$$\sigma_t^2 = w + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

$$E(\sigma_t^2) = w + \alpha E(\varepsilon_{t-1}^2) + \beta E(\sigma_{t-1}^2)$$

# GARCH(1,1)

$$\bar{\sigma}^2 = w + \alpha \bar{\sigma}^2 + \beta \bar{\sigma}^2$$

$$\bar{\sigma}^2 = \frac{w}{1-\alpha-\beta}$$

$$\mu_4 = \frac{3(1+\alpha+\beta)(1-\alpha-\beta)}{1-2\alpha\beta-3\alpha^2-\beta^2} > 3$$

# EGARCH

$$r_t = \mu + \varepsilon_t \text{ where } e_t \sim N(0,1)$$

$$\varepsilon_t = \sigma_t e_t$$

$$\ln(\sigma_t^2) = w + \alpha \left( \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) - \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2)$$

- The first component  $\left( \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) = |e_{t-1}| - \sqrt{\frac{2}{\pi}}$  is the absolute value of a normally distribution variable  $e_{t-1}$  minus its expectation.

# EGARCH

- The second component is  $\gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} = \gamma e_{t-1}$
- Notice that the two normal shocks behave differently.
- The first produces a symmetric rise in the log conditional variance.
- The second creates an asymmetric effect, in that, the log conditional variance rises following a negative shock.

# EGARCH

- More formally if  $e_{t-1} < 0$  then the log conditional variance rises by  $\alpha + \gamma$ .
- If  $e_{t-1} > 0$  then the log conditional variance rises by  $\alpha - \gamma$
- This produces the asymmetric volatility effect – volatility is higher during down market and lower during up market.

# Stochastic Volatility (SV)

➤ There is a variety of SV models.

➤ A popular one follows the dynamics

$$r_t = \mu_t + \sigma_t \varepsilon_t \text{ where } \varepsilon_t \sim N(0,1)$$

$$\ln(\sigma_t) = \gamma_0 + \gamma_1 \ln(\sigma_{t-1}) + \eta_t v_t \text{ where } v_t \sim N(0,1)$$

$$\text{cov}(v_t, \varepsilon_t) = 0$$

➤ Notice, that unlike ARCH, GARCH, and EGARCH, here the volatility itself has a stochastic component.

# Realized Volatility

- The realized volatility (RV) is a very tractable way to measure volatility.
- It essentially requires no parametric modeling approach.
- Suppose you observe daily observations within a trading month on the market portfolio.
- RV is the average of the squared daily returns within that month.



# Realized Volatility

- Of course, volatility varies on the monthly frequency but it is assumed to be constant within the days of that particular month.
- If you observe intra-day returns (available for large US firms) then daily RV is the sum of squared of five minute returns.
- You can use AR(1) to model log realized variance and then predict future values.

# Implied Volatility (IV)

- The B&S call Option price is given by

$$C(S, K, \sigma, r, T, \delta) = Se^{-\delta T} N(d_1) - Ke^{-rT} N(d_2)$$

- The put Option price is

$$P(S, K, \sigma, r, T, \delta) = Ke^{-rT} N(-d_2) - Se^{-\delta T} N(-d_1)$$

$$\text{Where } d_1 = \frac{\ln(S/K) + (r - \delta + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \text{ and } d_2 = d_1 - \sigma\sqrt{T}$$

# Implied Volatility (IV)

- In the traditional option pricing practice one inserts into the formula all the six parameters, i.e., the stock price, the strike price, the time to expiration, the cc risk-free rate, the cc dividend yield, and stock return volatility.
- IV is that volatility that if inserted into the B&S formula would yield the market price of the call or put option.
- As noted earlier, IV is not constant across maturities or across strike prices.

# Session #11 (part c): Stock Return Predictability, Model Selection, and Model Combination

# Return Predictability

- If log returns are IID – there is no way you can deliver better prediction for stock return than the current mean return.
- That is, if  $r_t = \mu + \varepsilon_t$  where  $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$

$$E[r_{t+1}|I_t] = \mu$$

$$Var[r_{t+1}|I_t] = \sigma^2$$

where  $r_t$  is the continuously compounded return and  $I_t$  is the set of information available at time  $t$ .

# Return Predictability

- Also note that the variance of a two-period return  $r_t + r_{t+1}$  is equal to

$$\text{Var}(r_t) + \text{Var}(r_{t+1}) + 2 \text{cov}(r_t, r_{t+1}) = 2\sigma^2$$

- That is, variance grows linearly with the investment horizon, while volatility grows in the rate square root.

# Variance Ratio Tests

- However, is it really the case?
- Perhaps stock returns are auto-correlated, or  $\text{cov}(r_t, r_{t+1}) \neq 0$

then:

$$\begin{aligned} VR_2 &= \frac{\text{Var}(r_t + r_{t+1})}{2\text{Var}(r_t)} = \frac{\text{Var}(r_t) + \text{Var}(r_{t+1}) + 2\text{cov}(r_t, r_{t+1})}{2\text{Var}(r_t)} \\ &= 1 + \frac{\text{cov}(r_t, r_{t+1})}{\sigma(r_t)\sigma(r_{t+1})} = 1 + \rho \end{aligned}$$



# Variance Ratio Tests

► Test:  $H_0: \rho = 0$

$H_1: \rho \neq 0$

► The test statistic is

$$\sqrt{T}(VR_2 - 1) \xrightarrow{d} N(0,1)$$

# Variance Ratio Tests

► More generally,

$$VR_g = \frac{\text{Var}(r_t + r_{t+1} + \dots + r_{t+g})}{(g+1)\text{Var}(r_t)} = 1 + 2 \sum_{s=1}^g \left(1 - \frac{s}{g+1}\right) \rho_s$$

$H_0: VR_g = 1$  no auto correlation

$H_1: \textit{Otherwise}$

# Variance Ratios

► Test statistic:

$$\sqrt{T}(VR_g - 1) \xrightarrow{d} N \left[ 0, \sum_{s=1}^{g-1} 4 \left( 1 - \frac{s}{g} \right)^2 \right]$$

e.g.  $g = 2$

$$\sqrt{T}(VR_2 - 1) \xrightarrow{d} N[0, 1]$$

$g = 3$

$$\sqrt{T}(VR_3 - 1) \xrightarrow{d} N \left[ 0, \frac{20}{9} \right]$$

# Predictive Variables

- In the previous specification, we used lagged returns to forecast future returns or future volatility.
- You can use a bunch of other predictive variables, such as:
  - The term spread.
  - The default spread.
  - Inflation.
  - The aggregate dividend yield

# Predictive Variables

- The aggregate book-to-market ratio.
- The market volatility.
- The market illiquidity

# Predictive Regressions

- To examine whether stock returns are predictable, we can run a predictive regression.
- This is the regression of future excess log or gross return on predictive variables.
- It is formulated as:

$$r_{t+1} = a + b_1 z_{1t} + b_2 z_{2t} + \dots + b_M z_{Mt} + \varepsilon_{t+1}$$

# Predictive Regressions

- To examine whether either of the macro variables can predict future returns test whether either of the slope coefficients is different from zero.
- Use the  $t$ -statistic or  $F$ -statistic for the regression  $R$ -squared.
- There is a small sample bias if (i) the predictive variables are highly persistent, (ii) the contemporaneous correlation between the predictive regression residual and the innovation of the predictor is high, or (iii) the sample is small.



# Long Horizon Predictive Regressions

$$r_{t+1,t+K} = \alpha + b'z_t + \varepsilon_{t+1,t+K}$$

- The dependent variable is the sum of log excess return over the investment horizon, which is  $K$  periods.
- Since the residuals are auto correlated compute the standard errors for the slope coefficient accounting for serial correlation and often for heteroscedasticity.
- For instance you can use the Newey-West correction.

# Newey-West Correction

► Rewriting the long horizon regression

$$r_{t+1,t+K} = x_t' \beta + \varepsilon_{t+1,t+K}$$

$$x_t' = [1, z_t']$$

$$\beta' = [a, b']$$

# Newey-West Correction

- The estimation error of the regression coefficient is represented by  $Var(\hat{\beta})$

$$Var(\hat{\beta}) = (X'X)^{-1}\hat{S}$$

- where  $\hat{S}$  is the Newey-West given by serially correlated adjusted estimator

$$\hat{S} = \sum_{j=-K}^K \frac{K - |j|}{K} \cdot \frac{1}{T} \sum_{t=j+1}^T \varepsilon_t \varepsilon_{t-j}$$

# Long Horizon Predictive Regressions

## Tradeoff:

- Higher  $K$  – better coverage of dependence.
- But we lose degrees of freedom.
- Feasible solution:

$$K \propto T^{\frac{1}{3}}$$

# Long Horizon Predictive Regressions

E.g.  $K=1$ :

$$j = -1, j = 0, j = 1$$

$$\hat{S} = \frac{1}{T} \sum_{t=1}^T \varepsilon_t^2$$

► Here we have no serial correlation.

# Long Horizon Predictive Regressions

E.g. K=2:

$$j = -2, j = -1, j = 0, j = 1, j = 2$$

$$\hat{S} = \frac{1}{2} \cdot \frac{1}{T} \sum_{t=1}^{T-1} \varepsilon_t \varepsilon_{t+1} + \frac{1}{T} \sum_{t=-1}^T \varepsilon_t^2 + \frac{1}{2} \cdot \frac{1}{T} \sum_{t=2}^T \varepsilon_t \varepsilon_{t-1}$$

$$= \frac{1}{T} \left[ \sum_{t=1}^T \varepsilon_t^2 + \sum_{t=1}^{T-1} \varepsilon_t \varepsilon_{t+1} \right]$$

# In the Presence of Heteroskedasticity

$$\text{Var}(\hat{\beta}) = \frac{1}{T} \left( \frac{1}{T} \sum_{t=1}^T x_t x_t' \right)^{-1} \hat{S} \left( \frac{1}{T} \sum_{t=1}^T x_t x_t' \right)^{-1}$$

$$\hat{S} = \sum_{j=-K}^K \frac{K - |j|}{K} \cdot \frac{1}{T} \sum_{t=1}^{T-K+1} \cdot [\varepsilon_t x_t x_{t-j}' \varepsilon_{t-j}]$$



# Out of Sample Predictability

➤ There is ample evidence of in-sample predictability, but little evidence of out-of-sample predictability.

➤ Consider the two specifications for the stock return evolution

$$M_1: r_t = a + bz_{t-1} + \varepsilon_t$$

$$M_2: r_t = \mu + \varepsilon_t$$

➤ Which one dominates? If  $M_1$  then there is predictability otherwise, there is no.

# Out of Sample Predictability

- One way to test predictability is to compute the out of sample  $R^2$ :

$$R_{OOS}^2 = 1 - \frac{\sum_{t=1}^T (r_t - \hat{r}_{t,1})^2}{\sum_{t=1}^T (r_t - \bar{r}_t)^2}$$

- Where  $\hat{r}_{t,1}$  is the return forecast assuming the presence of predictability, and  $\bar{r}_t$  is the sample mean (no predictability).
- Can compute the MSE (Mean Square Error) for both models.

# Model Selection

- When  $M$  variables are potential candidates for predicting stock returns there are  $2^M$  linear combinations of predictive models.
- In the extreme, the model that drops all predictors is the no-predictability or IID model.  
The one that retains all predictors is the all inclusive model.
- Which model to use?
- One idea (bad) is to implement model selection criteria.

# Model Selection Criteria

$$AIC = 2m - 2 \ln (L)$$

where  $L$  is the maximized value of the likelihood function.

$$BIC = m \ln (T) - 2 \ln (L)$$

$$\bar{R}^2 = 1 - (1 - R^2) \frac{T-1}{T-m-1} = R^2 - (1 - R^2) \frac{m}{T-m-1}$$

► Bayesian posterior probability

# Model Selection

- Notice that all criteria are a combination of goodness of fit and a penalty factor.
- You choose only one model and disregard all others.
- Model selection criteria have been shown to exhibit very poor out of sample predictive power.

# Model Combination

- The other approach is to combine models.
- Bayesian model averaging (BMA) computes posterior probabilities for each model then it uses the posterior probabilities as weights to compute the weighted model.
- There are more naive combinations.
- Such combination methods produce quite robust predictors not only in sample but also out of sample.