Topic 1: Introduction to Financial Data Analysis

Zongye Huang

International School of Economics and Management
Capital University of Economics and Business

Feb 20th, 2017

Outline

Outline

Financial Data Analysis I

- Instructors

Name:	Zongye Huang (黄宗晔)	Zheng Tian (田峥)
Email:	zongyeh@163.com	ztian_cueb@163.com
Office:	Chengming Building (诚明楼) 325	Angong Building (安工楼) 215

- Time and location

Odd weeks	Monday	13:30 – 15:20	Mingbian Building (明辨楼) 511
	Tuesday	15:40 – 17:30	Buoxue Building (博学楼) 316
Even weeks	Tuesday	15:40 – 17:30	Buoxue Building (博学楼) 316

- Office hours
  - Zongye Huang: Friday 13:30 – 14:30
  - Zheng Tian: Friday 13:30 – 14:30

You can also make an appointment by email.

Financial Data Analysis II

- Textbooks
  - *Primary:*
    - *"An Introduction to Analysis of Financial Data with R"*, 1st Edition, by Ruey S. Tsay.
  - *Reference:*
    - *"Analysis of Financial Time Series"*, 3rd Edition, by Ruey S. Tsay.
    - *"Applied Econometric Time Series"*, by Walter Enders
- Web Resources
  - Github:
    - https://isem-cueb-ztian.github.io/Financial-Data-Analysis-2017/
  - http://www.zongyehuang.com/financial-data-analysis
  - Baidu Yun
    - Link: https://pan.baidu.com/s/1i59MdQP Code: tqwp

Financial Data Analysis III

- Course outline

Instructors	Topics	Time
Zongye Huang	Introduction to financial data	Weeks 1
	Linear models for financial time series	Weeks 2 to 5
	Case studies of linear time series	Weeks 6 to 7
	Midterm exam	Week 8
Zheng Tian	Conditional heteroskedastic models	Weeks 9 to 12
	Value at Risk	Weeks 13 to 15
	Review and Q&A	Week 16
	Final exam	Week 17

Software: R

- R is the primary software we will use to analyze financial data. It is a free software available from http://www.r-project.org. It runs on many operating systems, including Linux, MacOS X, and Windows. You can download (or update to the latest version of) R at https://mirrors.tuna.tsinghua.edu.cn/CRAN/.
- The simplest way to install the program is to follow the online instructions and to use the default options. Because R is an open-source software, it contains hundreds of packages developed by researchers around the world for various statistical analyses.
- RStudio is the GUI (graphic user interface) of R. We will use it to edit R code and write empirical reports. Download it at https://www.rstudio.com/products/rstudio/download/.

Outline

Course Evaluation I

- Grade Composition

Component	%	Note
Homework	20%	
Project report	20%	
Midterm	30%	April 11
Final	30%	

- Teamwork
  - Students are encouraged to work in groups to complete their problem sets and project reports.
  - The formation of study groups is totally voluntary.
  - The size of each group should not exceed four students, and each student should only join one group.
  - Please send us the information of your study group no later than March 6th.

Course Evaluation II

- High resemblance of completed homework within each group is permitted. However, homework that is highly alike between groups will be treated as shirking, resulting in lower scores for all persons involved.
- Study groups are also course project groups. We want you to learn how to collaborate.
- A brief description of your contribution in the project can be attached with your report.

- Homework
    - Homework helps students understand fundamentals theoretical models and practice programming skills.
    - Every student should submit their own work.
    - Homework must be submitted on the due day announced in class, and email submission is accepted.

Course Evaluation III

- Course project
  - Course projects help student train research and writing skills as well as team working spirit. You can choose any topic and use any data set that are related to this course to complete a mini research project.
  - Course projects can be carried out individually or by study group. Each group should only submit one copy.
  - The final products of the project include:
    (1) a research report, (2) data and code used in the project, and (3) a documentation written in R Markdown that can be used to reproduce your results.
  - Complete explanations regarding research reports and documentation will be given in class.

Course Evaluation IV

- The mid-term exam
  - The mid-term exam is tentatively scheduled on **April 11**, which will cover all contents of the first part.
  - It will be a closed-book test.
  - If you miss the mid-term exam, a make-up test can be arranged. You must notify me of your absence in advance with a valid excuse.

Course Evaluation V

- The final exam
  - The final exam is in Week 17, covering all content that Prof. Tian teaches.
  - It will also be a closed-book. But you are allowed to bring a one-sided "cheat sheet", on which you can write down some notes that help you remember some important definitions and formula. You are allowed to write on only one side on the cheat sheet.
  - The time and location are to be arranged and announced by the university.
  - The make-up test will follow the rule of the university.

Outline

One-Period Simple Return

- Holding the asset for one period from date $t-1$ to date t would result in a simple gross return

$$1+R_t = \frac{P_t}{P_{t-1}}.$$

- The corresponding one-period simple net return or simple return is

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

Multi-Period Simple Return I

- Holding the asset for k periods from date $t-k$ to date t would result in a k-period simple gross return

$$
\begin{aligned}
1+R_t[k] &= \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \cdot \frac{P_{t-1}}{P_{t-2}} \cdot \ldots \cdot \frac{P_{t-k+1}}{P_{t-k}} \\
&= (1+R_t)(1+R_{t-1}) \cdot \ldots \cdot (1+R_{t-k+1}) \\
&= \prod_{j=0}^{k-1} (1+R_{t-j})
\end{aligned}
$$

the k-period simple gross return is just the product of the k one-period simple gross returns involved. This is called a **compound return**.

Multi-Period Simple Return II

- The corresponding k-period simple net return is

$$R_t[k] = \frac{P_t}{P_{t-k}} - 1 = \frac{P_t - P_{t-k}}{P_{t-k}}.$$

- In practice, the actual time interval is important in discussing and comparing returns (e.g., monthly return or annual return). If the time interval is not given, then it is implicitly assumed to be one year.
- If the asset was held for k years, then the annualized (average) return is defined as

$$\text{Annualized}\{R_t[k]\} = \left[\prod_{j=0}^{k-1}(1 + R_{t-j})\right]^{\frac{1}{k}} - 1.$$

Multi-Period Simple Return III

- Alternatively, this geometric mean involving k one-period simple gross returns can be computed by

$$\text{Annualized}\{R_t[k]\} = \exp\left[\frac{1}{k}\sum_{j=0}^{k-1}\ln(1+R_{t-j})\right] - 1,$$

because it is easier to compute arithmetic average than geometric mean.

- Since the one-period returns tend to be small, we can use a first-order Taylor expansion to approximate the annualized return

$$\text{Annualized}\{R_t[k]\} \approx \frac{1}{k}\sum_{j=0}^{k-1} R_{t-j}.$$

Continuously Compound Return

- The natural logarithm of the simple gross return of an asset is called the continuously compounded return or log return:

$$r_t = \ln(1+R_t) = \ln\left(\frac{P_t}{P_{t-1}}\right) = p_t - p_{t-1}$$

  where $p_t = \ln P_t$.

- The continuously compounded multiperiod return is simply the sum of continuously compounded one-period returns involved

$$\begin{aligned} r_t[k] &= \ln(1+R_t[k]) = \ln\left[(1+R_t)(1+R_{t-1})\cdot\ldots\cdot(1+R_{t-k+1})\right] \\ &= r_t + r_{t-1} + \ldots + r_{t-k+1} \end{aligned}$$

Dividend Payment

- If an asset pays dividends periodically, we must modify the definitions of asset returns. Let D_t be the dividend payment of an asset between dates $t-1$ and t, and P_t be the price of the asset at the end of period t.
- Since dividend is not included in P_t, the simple net return and continuously compounded return at time t become

$$
\begin{aligned}
1+R_t &= \frac{P_t+D_t}{P_{t-1}}, \\
r_t &= \ln(P_t+D_t)-\ln(P_{t-1}).
\end{aligned}
$$

Excess Return

- Excess return of an asset at time t is the difference between the asset's return and the return on some reference asset.
    - Which reference asset?
        - Bearing very little risk.
        - Price is determined in the market.
        - Trade at high frequency.
    - The reference asset is often taken to be riskless such as a short-term U.S. Treasury bill return.
- The simple excess return and log excess return of an asset are then defined as

$$Z_t = R_t - R_{0t}.$$

Example

1. If the monthly log return of an asset is 4.46%, what is the monthly simple return?
   The corresponding monthly simple return is

$$100[\exp(\frac{4.46}{100}) - 1] = 4.56$$

2. If the monthly log returns of the asset within a quarter are 4.46%, $-$ 7.34% , and 10.77%, respectively, then what is the quarterly log return of the asset?
   The quarterly log return of the asset is

$$(4.46 - 7.34 + 10.77) = 7.89$$

Outline

Bond Yields I

- Bonds are a financial instrument that will pay the face value (or par value) to its holder at the time of maturity.
  - Some bonds also pay interest periodically referring to as coupon payment.
  - Zero-coupon bonds do not pay periodic interest.
- **Bond yield** is the return an investor will receive by holding a bond to maturity. The common ones are the current yield and yield to maturity (YTM).

Bond Yields II

- **Current Yield**
  - The ratio of the interest rate payable on a bond to the actual market price of the bond, stated as a percentage (Securities Industry and Financial Markets Association).

$$\text{Current yield} = \frac{\text{Annual interest paid in dollars}}{\text{Market price of the bond}} \times 100\%.$$

  - For example, a bond with a current market price of par (\$1,000) that pays eighty dollars (\$80) per year in interest would have a current yield of eight percent.
- If an investor paid \$90 for a bond with **face value** of \$100, also known as **par value**, and the bond paid a coupon rate of 5% per annum, then the current yield of the bond is $c_t = (0.05 \times 100)/90 \times 100$.

Bond Yields III

- For *zero-coupon bonds*, the yield is

$$\text{Current yield} = \left(\frac{\text{Face value}}{\text{Purchase price}}\right)^{1/k} - 1$$

  where k denotes time to maturity in years.

  - If an investor purchased a zero-coupon bond with face value \$100 for \$95 and the bond will mature in 6 months (182 days), the yield is $c_t = (100/95)^2 - 1 = 10.8\%$.

- Discount note (Securities Industry and Financial Markets Association)
  - Short-term obligations issued at a discount from face value, with maturities ranging from one to 360 days. Discount notes have no periodic interest payments; the investor receives the note's face value at maturity.
  - For example, a one year, \$1,000 face value discount note purchased at issue at a price of \$950, would yield \$50 or 5.26 percent (\$50/\$950).

Bond Yields IV

- **Yield to Maturity (YTM)** is the yield obtained by equating the bond price to the present value of all future payments.
  - Suppose that the bond holder will receive k payments between purchase and maturity. y is the YTM and P is the price of the bond, C_i is the ith cash flow of coupon payment, F denotes the face value

$$P = \frac{C_1}{1+y} + \frac{C_2}{(1+y)^2} + \ldots + \frac{C_k + F}{(1+y)^k}.$$

  - Suppose that the coupon rate is α per annum, the number of payments is m per year, and the time to maturity is n years.

Bond Yields V

- Cash flow of coupon payment is $F\alpha/m$, and the number of payments is $k = mn$. The bond price and YTM is

$$
\begin{aligned}
P &= \frac{F\alpha}{m}\left[\frac{1}{1+y} + \frac{1}{(1+y)^2} + \ldots + \frac{1}{(1+y)^k}\right] + \frac{F}{(1+y)^k} \\
&= \frac{F\alpha}{my}\left[1 - \frac{1}{(1+y)^k}\right] + \frac{F}{(1+y)^k}
\end{aligned}
$$

The table below gives some results between bond price and YTM assuming that $F = \$100$, coupon rate is 5% per annum payable semiannually, and time to maturity is 3 years.

Yield to Maturity (%)	Semiannual Rate (%)	Bond Price ($)
6	3.0	97.29
7	3.5	94.67
8	4.0	92.14
9	4.5	89.68
10	5.0	87.31

U.S. Government Bonds I

The U.S. Government issues various bonds to finance its debts, including Treasury bills, Treasury notes, and Treasury bonds.

- **Treasury bills (T-Bills)** mature in one year or less. They do not pay interest prior to maturity and are sold at a discount of the face value (or par value) to create a positive YTM. The commonly used maturities are 28 days (1 month), 91 days (3 months), 182 days (6 months), and 364 days (1 year). The minimum purchase is \$100. The discount yield of T-Bills is

$$\text{Discount yield } (\%) = \frac{F - P}{F} \times \frac{360}{\text{Days till maturity}} \times 100(\%)$$

U.S. Government Bonds II

- **Treasury notes (T-Notes)** mature in 1–10 years. They have a coupon payment every 6 months and face value of \$1000. These notes are quoted on the secondary market at percentage of face value in thirty-seconds of a point. For example, a quote 95:08 on a note indicates that it is trading at a discount $\$(95 + 8/32) \times 1000 = \$952.5$.
    - The 10-year Treasury note has become the security most frequently quoted when discussing the U.S. government bond market.

U.S. Government Bonds III

- **Treasury bonds (T-Bonds)** have longer maturities, ranging from 20 to 30 years. They have a coupon payment every 6 months and are commonly issued with maturities 30 years. The 30-year bonds were suspended for a 4-year and 6- month period starting October 31, 2001, but they were reintroduced in February 2006 and are now issued quarterly.

Outline

Distributional Properties of Returns

- To study asset returns, it is best to begin with their distributional properties.
- The objective here is to understand the behavior of the returns across assets and over time.
- Consider a collection of N assets held for T time periods, say, $t = 1, ..., T$.
  - For each asset i, let r_{it} be its log return at time t. The log returns under study are $r_{it}; i = 1, ..., N; t = 1, ..., T$.
  - One can also consider the simple returns $R_{it}; i = 1, ..., N; t = 1, ..., T$ and the log excess returns $z_{it}; i = 1, ..., N; t = 1, ..., T$.

Review Statistical Distributions I

- Let R^k be the k-dimensional Euclidean space. A point in R^k is denoted by $x \in R^k$. Consider two random vectors $X = (X_1, ..., X_k)$ and $Y = (Y_1, ..., Y_q)$.
- Let $P(X \in A, Y \in B)$ be the probability that X is in the subspace $A \subset R^k$ and Y is in the subspace $B \subset R^q$.
- For most of the cases considered in this book, both random vectors are assumed to be continuous.

Review Statistical Distributions II

Joint Distribution

The function

$$F_{X,Y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) = P(\boldsymbol{X} \leq \boldsymbol{x}, \boldsymbol{Y} \leq \boldsymbol{y}; \boldsymbol{\theta}),$$

where $\boldsymbol{x} \in R^p$, $\boldsymbol{y} \in R^q$, and the inequality $\leq$ is a component-by-component operation, is a joint distribution function of $\boldsymbol{X}$ and $\boldsymbol{Y}$ with parameter $\boldsymbol{\theta}$. Behavior of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is characterized by $F_{X,Y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})$. If the joint probability density function $f_{x,y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})$ of $\boldsymbol{X}$ and $\boldsymbol{Y}$ exists, then

$$F_{X,Y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) = \int_{-\infty}^{\boldsymbol{x}} \int_{-\infty}^{\boldsymbol{y}} f_{x,y}(\boldsymbol{w}, \boldsymbol{z}; \boldsymbol{\theta}) \, d\boldsymbol{z} \, d\boldsymbol{w}.$$

In this case, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are continuous random vectors.

Review Statistical Distributions III

Marginal Distribution

The marginal distribution of $\boldsymbol{X}$ is given by

$$F_X(\boldsymbol{x};\boldsymbol{\theta}) = F_{X,Y}(\boldsymbol{x}, \infty, \cdots, \infty; \boldsymbol{\theta}).$$

Thus, the marginal distribution of $\boldsymbol{X}$ is obtained by integrating out $\boldsymbol{Y}$. A similar definition applies to the marginal distribution of $\boldsymbol{Y}$.

If $k = 1$, X is a scalar random variable and the distribution function becomes

$$F_X(x) = P(X \leq x; \boldsymbol{\theta}),$$

Review Statistical Distributions IV

Conditional Distribution. The conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y} \leq \boldsymbol{y}$ is given by

$$F_{X|Y\leq y}(\boldsymbol{x};\boldsymbol{\theta}) = \frac{P(\boldsymbol{X} \leq \boldsymbol{x}, \boldsymbol{Y} \leq \boldsymbol{y};\boldsymbol{\theta})}{P(\boldsymbol{Y} \leq \boldsymbol{y};\boldsymbol{\theta})}.$$

If the probability density functions involved exist, then the conditional density of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$ is

$$f_{x|y}(\boldsymbol{x};\boldsymbol{\theta}) = \frac{f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})}{f_y(\boldsymbol{y};\boldsymbol{\theta})}, \tag{1.8}$$

where the marginal density function $f_y(\boldsymbol{y};\boldsymbol{\theta})$ is obtained by

$$f_y(\boldsymbol{y};\boldsymbol{\theta}) = \int_{-\infty}^{\infty} f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})\mathrm{d}\boldsymbol{x}.$$

Review Statistical Distributions V

From Equation (1.8), the relation among joint, marginal, and conditional distributions is

$$f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta}) = f_{x|y}(\boldsymbol{x};\boldsymbol{\theta}) \times f_y(\boldsymbol{y};\boldsymbol{\theta}). \tag{1.9}$$

This identity is used extensively in time series analysis (e.g., in maximum likelihood estimation). Finally, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent random vectors if and only if $f_{x|y}(\boldsymbol{x};\boldsymbol{\theta}) = f_x(\boldsymbol{x};\boldsymbol{\theta})$. In this case, $f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta}) = f_x(\boldsymbol{x};\boldsymbol{\theta})f_y(\boldsymbol{y};\boldsymbol{\theta})$.

Moments of a Random Variable I

Moments of a Random Variable. The ℓth moment of a continuous random variable X is defined as

$$m'_\ell = E(X^\ell) = \int_{-\infty}^{\infty} x^\ell f(x)\mathrm{d}x,$$

where E stands for expectation and $f(x)$ is the probability density function of X. The first moment is called the *mean* or *expectation* of X. It measures the central location of the distribution. We denote the mean of X by μ_x. For an asset, an interesting question is whether the mean of its return is zero. In other words, we often consider the hypothesis testing $H_0 : \mu_x = 0$ versus $H_a : \mu \neq 0$ or $H_0 : \mu_x \leq 0$ versus $H_a : \mu_x > 0$.

The ℓth central moment of X is defined as

$$m_\ell = E[(X - \mu_x)^\ell] = \int_{-\infty}^{\infty} (x - \mu_x)^\ell f(x)\mathrm{d}x$$

In finance, the first fourth moments of a random variable are used to describe the behavior of asset returns.

Moments of a Random Variable II

- The **second** central moment,
  - σ_x^2, the variance, measures the variability of X.

$$\begin{aligned}\sigma_x^2 &= E\left[X-\mu_x\right]^2 = E\left[X^2 - 2\mu_x X + \mu_x^2\right] \\ &= E\left[X^2\right] - \mu_x^2\end{aligned}$$

  - The positive square root, σ_x, of variance is the standard deviation of X.
  - For asset returns, variance (or standard deviation) is a measure of uncertainty, used as a risk measure.

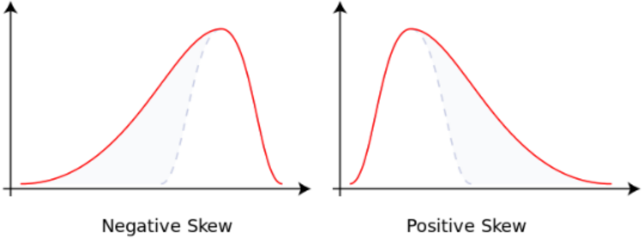
The first two moments of a random variable uniquely determine a normal distribution. For other distributions, higher order moments are also of interest.

Moments of a Random Variable III

- The **third** central moment measures the symmetry of X with respect to its mean.
    - Standardized third moment is called **skewness**.

$$S(x) = E\left[\frac{(X-\mu_x)^3}{\sigma_x^3}\right].$$

Moments of a Random Variable IV



- Negative skew: The left tail is longer; the mass of the distribution is concentrated on the right of the figure, left-skewed, left-tailed, or skewed to the left.
- Positive skew: The right tail is longer; the mass of the distribution is concentrated on the left of the figure, right-skewed, right-tailed, or skewed to the right.

Moments of a Random Variable V

- The **fourth** central moment measures the tail behavior of X.
    - Normalized third moment is called **kurtosis**. which summarizes the extent of tail thickness.

$$K(x) = E\left[\frac{(X-\mu_x)^4}{\sigma_x^4}\right].$$

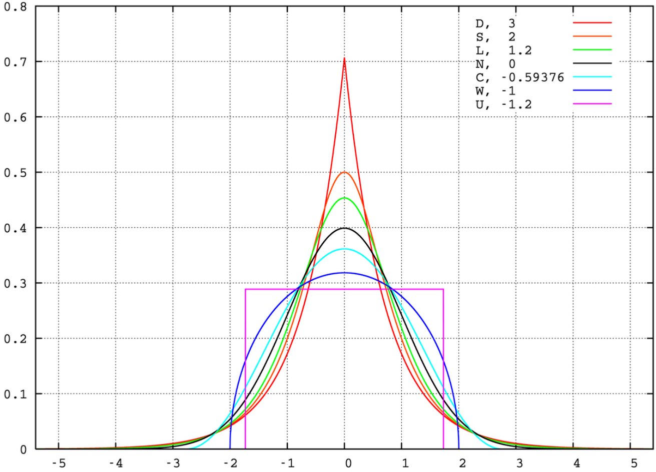
    - The quantity $K(x)-3$ is called the **excess kurtosis**, because $K(x)=3$ for a normal distribution.
        - The excess kurtosis of a normal random variable is zero.
- A distribution with positive excess kurtosis is said to have heavy tails.
    - It puts more mass on the tails of its support than a normal distribution does, meaning that a random sample from such a distribution tends to contain more extreme values.
    - Called **leptokurtic,** super-Gaussian.

Moments of a Random Variable VI

- Student's t-distribution, exponential distribution, Poisson distribution.

- A distribution with negative excess kurtosis has short tails
  - Called **platykurtic**, sub-Gaussian.
  - Continuous or discrete uniform distributions

Moments of a Random Variable VII



Probability density functions for selected distributions with mean 0, variance 1 and different excess kurtosis.

Estimate Sample Moments I

Moments of a random variable can be estimated by their sample counterparts. Let $x_1, ..., x_T$ be a random sample of X with T observations.

- The sample mean is

$$\hat{\mu}_x = \frac{1}{T} \sum_{t=1}^{T} x_t.$$

- The sample variance is

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

Estimate Sample Moments II

- The sample skewness is

$$\hat{S}(x) = \frac{1}{(T-1)\hat{\sigma_x}^3} \sum_{t=1}^{T} (x_t - \hat{\mu_x})^3 .$$

- The sample kurtosis is

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma_x}^4} \sum_{t=1}^{T} (x_t - \hat{\mu_x})^4 .$$

Estimate Sample Moments III

Under rather weak conditions, the sample mean $\hat{\mu}_x$ is a consistent estimate of μ_x, meaning that $\hat{\mu}_x$ converges to μ_x as $T \to \infty$. More specifically, we have $\hat{\mu}_x \sim N(\mu_x, \sigma_x^2/T)$ for a sufficiently large T. This result is often used to test any hypothesis about μ_x. For instance, consider $H_0 : \mu_x = 0$ versus $H_a : \mu_x \neq 0$. The test statistic is

$$t = \frac{\sqrt{T}\hat{\mu}_x}{\hat{\sigma}_x},$$

which follows a Student's-t distribution with $T-1$ degrees of freedom. For a sufficiently large T, the test statistic approaches a standard normal distribution. The decision rule is then to reject H_0 at the $100\alpha\%$ level if $|t| > Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard normal distribution. Most statistical packages now provide p-value for each test statistic. The decision rule is then to reject H_0 at the $100\alpha\%$ level if the p-value is less than α.

- If X is a normal random variable, then $\hat{S}(x)$ and $\hat{K}(x) - 3$ are distributed asymptotically as normal with zero mean and variances $6/T$ and $24/T$, respectively.

Estimate Sample Moments IV

- This can be used to test the normality of asset returns. Given an asset return series $r_1, \ldots, r_T$, to test the skewness of the returns, we consider the null hypothesis $H_o : S(r) = 0$ versus the alternative hypothesis $H_a : S(r) \neq 0$.
  - The t-ratio statistic of the sample skewness is

$$t = \frac{\hat{S}(x)}{\sqrt{6/T}}.$$

  - The decision rule is to reject the null hypothesis at the $100\alpha\%$ significance level, if $|t| > Z_{1-\alpha/2}$.
- To test the excess kurtosis of the return series using the hypotheses $H_o : K(r) - 3 = 0$ versus $H_a : K(r) - 3 \neq 0$.
  - The test statistic is

$$t = \frac{\hat{K}(x) - 3}{\sqrt{24/T}}.$$

Estimate Sample Moments V

- Jarque and Bera (1987) combine the two prior tests to test for the normality of r_t

$$JB = \frac{\left(\hat{S}(x)\right)^2}{6/T} + \frac{\left(\hat{K}(x) - 3\right)^2}{24/T}$$

which is asymptotically distributed as a chi-squared random variable with 2 degrees of freedom.

Something left from last week

1. The "Volume" of a stock: number of stocks trades.
2. TNX: the CBOE 10-Year Treasury Note (TNX) is based on 10 times the yield-to-maturity on the most recently auctioned 10-year Treasury note.
3. How to get dividend data? Search for "dividend quantmod". Do not use Baidu.
   http://www.quantmod.com/documentation/getDividends.html
4. Why people trade in the market?

In my opinion:

- Different risk exposure.
- Have access to information.
- Response to random shock.
- Over shooting.
  Herd behavior.

Outline

Some Statistical Distributions I

Several statistical distributions have been proposed in the literature for the marginal distributions of asset returns, including

- normal distribution,
- lognormal distribution,
- stable distribution,
- scale mixture of normal distributions.

Some Statistical Distributions II

Normal Distribution

A traditional assumption made in financial study is that the simple returns $\{R_{it}|t = 1, \ldots, T\}$ are independently and identically distributed as normal with fixed mean and variance. This assumption makes statistical properties of asset returns tractable. But it encounters several difficulties. First, the lower bound of a simple return is -1. Yet the normal distribution may assume any value in the real line and, hence, has no lower bound. Second, if R_{it} is normally distributed, then the multiperiod simple return $R_{it}[k]$ is not normally distributed because it is a product of one-period returns. Third, the normality assumption is not supported by many empirical asset returns, which tend to have a positive excess kurtosis.

Some Statistical Distributions III

Lognormal Distribution

Another commonly used assumption is that the log returns r_t of an asset are independent and identically distributed (iid) as normal with mean μ and variance σ^2. The simple returns are then iid lognormal random variables with mean and variance given by

$$E(R_t) = \exp\left(\mu + \frac{\sigma^2}{2}\right) - 1, \quad \mathrm{Var}(R_t) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]. \quad (1.17)$$

These two equations are useful in studying asset returns (e.g., in forecasting using models built for log returns). Alternatively, let m_1 and m_2 be the mean and variance of the simple return R_t, which is lognormally distributed. Then the mean and

Some Statistical Distributions IV

variance of the corresponding log return r_t are

$$E(r_t) = \ln\left[\frac{m_1 + 1}{\sqrt{1 + m_2/(1 + m_1)^2}}\right], \qquad \mathrm{Var}(r_t) = \ln\left[1 + \frac{m_2}{(1 + m_1)^2}\right].$$

Because the sum of a finite number of iid normal random variables is normal, $r_t[k]$ is also normally distributed under the normal assumption for $\{r_t\}$. In addition, there is no lower bound for r_t, and the lower bound for R_t is satisfied using $1 + R_t = \exp(r_t)$. However, the lognormal assumption is not consistent with all the properties of historical stock returns. In particular, many stock returns exhibit a positive excess kurtosis.

Some Statistical Distributions V

Stable Distribution

The stable distributions are a natural generalization of normal in that they are stable under addition, which meets the need of continuously compounded returns r_t. Furthermore, stable distributions are capable of capturing excess kurtosis shown by historical stock returns. However, nonnormal stable distributions do not have a finite variance, which is in conflict with most finance theories. In addition, statistical modeling using nonnormal stable distributions is difficult. An example of nonnormal stable distributions is the Cauchy distribution, which is symmetric with respect to its median but has infinite variance.

Some Statistical Distributions VI

Scale Mixture of Normal Distributions

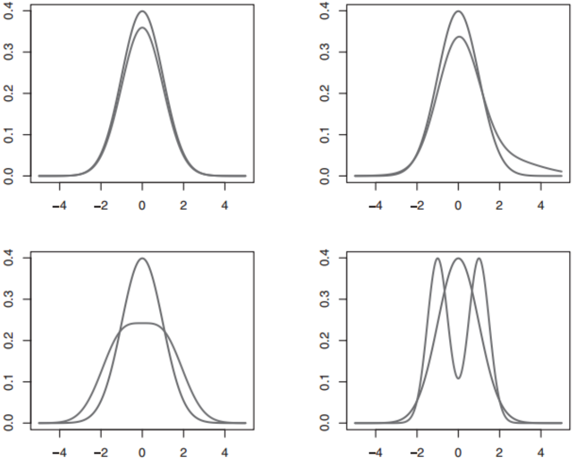
Recent studies of stock returns tend to use scale mixture or finite mixture of normal distributions. Under the assumption of scale mixture of normal distributions, the log return r_t is normally distributed with mean μ and variance σ^2 [i.e., $r_t \sim N(\mu, \sigma^2)$]. However, σ^2 is a random variable that follows a positive distribution (e.g., σ^{-2} follows a gamma distribution). An example of finite mixture of normal distributions is

$$r_t \sim (1 - X)N(\mu, \sigma_1^2) + XN(\mu, \sigma_2^2),$$

where X is a Bernoulli random variable such that $P(X = 1) = \alpha$ and $P(X = 0) = 1 - \alpha$ with $0 < \alpha < 1$, σ_1^2 is small, and σ_2^2 is relatively large. For instance, with $\alpha = 0.05$, the finite mixture says that 95% of the returns follow $N(\mu, \sigma_1^2)$ and 5% follow $N(\mu, \sigma_2^2)$. The large value of σ_2^2 enables the mixture to put more mass at the tails of its distribution. The low percentage of returns that are from $N(\mu, \sigma_2^2)$ says that the majority of the returns follow a simple normal distribution. Advantages of mixtures of normal include that they maintain the tractability of normal, have finite higher order moments, and can capture the excess kurtosis. Yet it is hard to estimate the mixture parameters (e.g., the α in the finite-mixture case).

Some Statistical Distributions VII

Mixtures of Normals



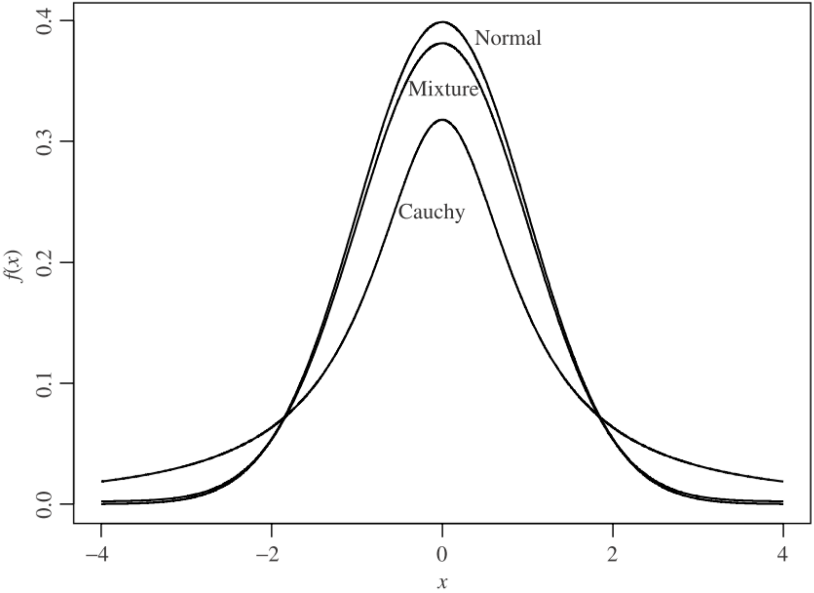
Some Statistical Distributions VIII

Figure 1.17 shows the probability density functions of a finite mixture of normal, Cauchy, and standard normal random variable. The finite mixture of normal is $(1 - X)N(0, 1) + X \times N(0, 16)$ with X being Bernoulli such that $P(X = 1) = 0.05$, and the density function of Cauchy is

$$f(x) = \frac{1}{\pi(1 + x^2)}, \qquad -\infty < x < \infty.$$

It is seen that the Cauchy distribution has fatter tails than the finite mixture of normal, which, in turn, has fatter tails than the standard normal.

Some Statistical Distributions IX



Homework #1 I

1. Form a study group with a group leader.
2. Install RStudio and packages.
3. Each member select one type of asset, such as stock, commodity, option, index, etc.
   3.1 Describe the economic background of the asset.
   3.2 Download the data. Quantmod may help you.
   3.3 Compute the sample mean, standard deviation, skewness, excess kurtosis, minimum, and maximum of each simple return series.
   3.4 Transform the simple returns to log returns. Compute the sample mean, standard deviation, skewness, excess kurtosis, minimum, and maximum of each log return series.
   3.5 Plot the data in at least two figures, such as histogram, ohlc plot, moving average.
   3.6 Test the null hypothesis that the mean of the log returns is zero. Use 5% significance level to draw your conclusion.

Homework #1 II

3.7 Conduct test for skewness, excess kurtosis, and normality.

4. There is also an optional question for the group.
   What is the excess Kurtosis of a scale mixture Normal Distribution?
   - Define a scale mixture Normal Distribution
   - Draw a random sample
   - Calculate the sample Kurtosis
   - Conduct the excess Kurtosis test
5. Group leader collect reports and submit them to zongyeh@163.com before **March 14**.