Computational Finance and Data Analysis for Financial Engineering

Zhipeng Liao

UCLA

Version 1.0

Introduction

- **Instructor:** Zhipeng Liao
  - Bunche Hall, 8379
  - E-mail: zhipeng.liao at econ.ucla.edu
- **Time and Location:** MW 12:30pm – 1:45pm, HUMANTS A65
- **Office Hours:** W 2:00 pm – 3:00 pm
  - or by appointment

Textbooks

- **Main:** *An Introduction to Computational Finance and Financial Econometrics*, by Eric Zivot
  - manuscript in preparation. Pdf files will be provided through course webpage
  - Class slides and notes
- **Supplementary**
  - *Statistics and Data Analysis for Financial Engineering*, by David Ruppert, Springer-Verlag
  - *The Elements of Financial Econometrics*, by Jianqing Fan and Qiwei Yao, Cambridge University Press

Evaluation

- **Homework** (approximately bi-weekly), 20%.
- **Mid-term Exam** (May 7th, 2018 in class), 30%
- **Final Exam** (June 11th, 2018), 50%
- **Please double check the final exam date and location before the exam**

Homework Problem Sets

- Problem sets will be assigned between Mon - Wed (bi-weekly, approximately)
- Due by next Monday 12:30 pm before class
- Drop your homework in an assigned box before class
  - Any late submission will not be accepted.
  - Make a report, using word processor (e.g., MS word), for the statistical output.
- Use our discussion page for your questions
  - discuss problem sets with classmates
  - submit your own answers
- Around **60%** of midterm and final questions will be based on the homework problems. Numbers will be changed.

Mid-term Exam

- The mid-term exam is scheduled on **Monday, May 7, 2018**
- The mid-term will cover everything discussed during the **April 2, 2018 – May 2, 2018** window
- A practice mid-term exam will be provided
- The mid-term exam will be the same style as the practice mid-term exam

Final Exam

- The final exam is scheduled on **Monday, June 11, 2018**
- Please **double check** the date and the location of the final before the exam
- The final exam will cover everything discussed this quarter, including the material covered in the midterm exam
- A practice final exam will be provided
- The final exam will be the same style as the practice final exam.

Grading

- Final grading based on the followings

Assignments	20%
Mid-term Exam	30%
Final Exam	50%

- Your score of the assignments will be divided by the maximum possible score, and then multiplied by 100. Call it **H**.
- Your midterm score will be divided by the maximum possible midterm score, and then multiplied by 100. Call it **M**.
- Your final score will be divided by the maximum possible final score, and then multiplied by 100. Call it **U**.
- Your weighted average **W** is calculated by the formula $\mathbf{W} = \mathbf{H} \times 0.2 + \mathbf{M} \times 0.3 + \mathbf{U} \times 0.5$.

Grading Algorithm 1

- If $95 \leq \mathbf{W}$, you will get A+. If $85 \leq \mathbf{W} < 95$, you will get A. If $80 \leq \mathbf{W} < 85$, you will get A-;
- If $75 \leq \mathbf{W} < 80$, you will get B+. If $70 \leq \mathbf{W} < 75$, you will get B. If $65 \leq \mathbf{W} < 70$, you will get B-;
- If $60 \leq \mathbf{W} < 65$, you will get C+. If $55 \leq \mathbf{W} < 60$, you will get C. If $50 \leq \mathbf{W} < 55$, you will get C-;
- If $\mathbf{W} < 50$, you will get F.

Grading Algorithm 2

- Your **W** will determine your rank **R** in the class (see the course syllabus for how your rank is determined);
- Let n1, n2, n3, n4, n5, n6, n7, n8 and n9 denote the smallest integers larger than or equal to $0.05 \times n$, $0.20 \times n$, $0.30 \times n$, $0.40 \times n$, $0.60 \times n$, $0.70 \times n$, $0.80 \times n$, $0.85 \times n$ and $0.90 \times n$ respectively.
- If $\mathbf{R} \leq$ n1, you will get A+. If n1 $< \mathbf{R} \leq$ n2, you will get A. If n2 $< \mathbf{R} \leq$ n3, you will get A-.
- If n3 $< \mathbf{R} \leq$ n4, you will get B+. If n4 $< \mathbf{R} \leq$ n5, you will get B. If n5 $< \mathbf{R} \leq$ n6, you will get B-.
- If n6 $< \mathbf{R} \leq$ n7, you will get C+. If n7 $< \mathbf{R} \leq$ n8, you will get C. If n8 $< \mathbf{R} \leq$ n9, you will get C-.
- If n9 $< \mathbf{R}$, you will get F.

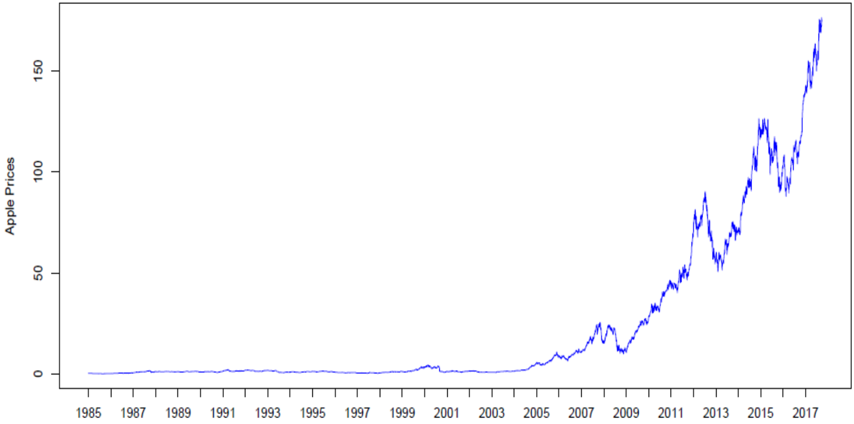
Grading

- Your **final grade** will be the maximum of the two grades determined in algorithms 1 and 2 above
- The secrete of getting a decent grade in this course
  - you should really understand the homework problems
  - you get $\mathbf{H}\times 0.2 \approx 20$ if you finish and submit all the homework assignments
  - you get $\mathbf{M}\times 0.3 \approx 18$ and $\mathbf{U}\times 0.5 \approx 30$ if you understand the homework problems and solve similar problems correctly in the exams
  - you get $\mathbf{W} \approx 68$ and hence a grade at least (around) B- based on grading algorithm 1
  - "$\mathbf{a} \approx \mathbf{b}$" means number **a** approximately equals to number **b**
- The secrete of getting a good grade in this course
  - you should also really understand all the material discussed in class

Course Description

- The price of risky asset usually fluctuates over time
- This gives us opportunities of making money by buying or selling these assets
- If we make an investment, the most important thing is how much money we will make
- The first topic in this course is on **calculation of the financial returns**.

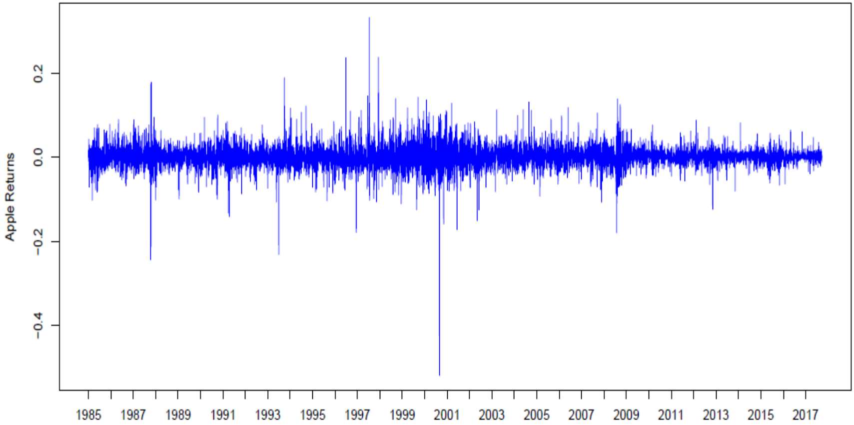
Daily Prices of Apple Shares (01/01/1985 – 12/26/2017)



- Buy 1 share at price P_t and sell it at price P_{t+1}
- Cost: P_t
- Revenue: P_{t+1}
- Profit: $P_{t+1} - P_t$
- The simple return

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

Daily Returns of Apple Shares (01/01/1985 – 12/26/2017)



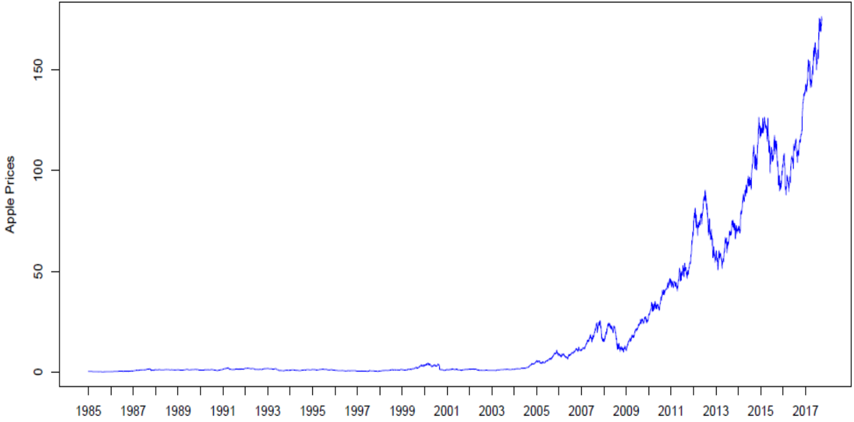
Course Description

- The financial return is a random variable, e.g., it may be positive or negative
- The randomness of the financial return leads to the risk of an investment
- Since risk depends upon the probability distribution of a return, **probability** and **statistics** are fundamental tools for finance
- **Probability** is needed for risk calculations (which assume we know the probability distribution of a return)
- **Statistics** is needed to estimate parameters of the probability distribution of a return, such as its mean and variance, and to test various financial theories
- The second topic in this course is on **review of concepts in probability and statistics**

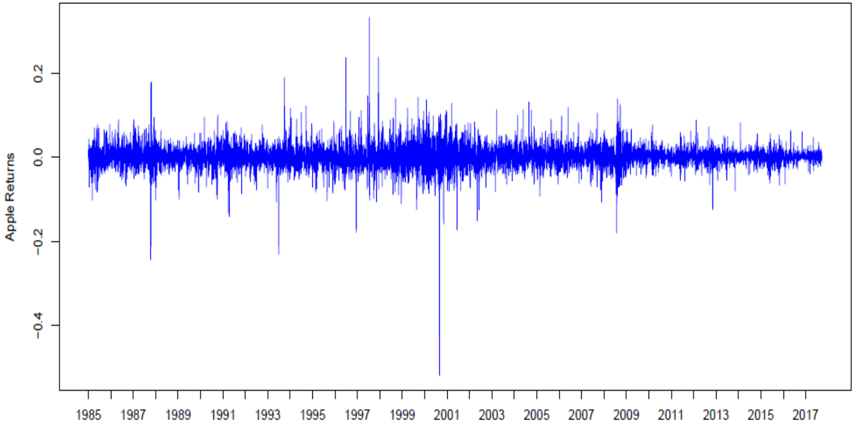
Course Description

- The simple return at any period t, say R_t, is a random variable
- The collection of financial returns at different periods, i.e., $\{\ldots, R_t, R_{t+1}, R_{t+2}, \ldots\}$, is a set of random variables
- Such a set of random variables is called a **time series**
- The financial data we observe in reality is a snapshot of the time series $\{\ldots, R_t, R_{t+1}, R_{t+2}, \ldots\}$ at a given window
- The **time series analysis** is on the probability and statistics of a set of random variables such as financial returns
- The third topic in this course is on **the time series econometrics**

Daily Prices of Apple Shares (01/01/1985 – 12/26/2017)



Daily Returns of Apple Shares (01/01/1985 – 12/26/2017)



- Covariance between two random variables X and Y: $Cov(X, Y)$
- The correlation between X and Y:

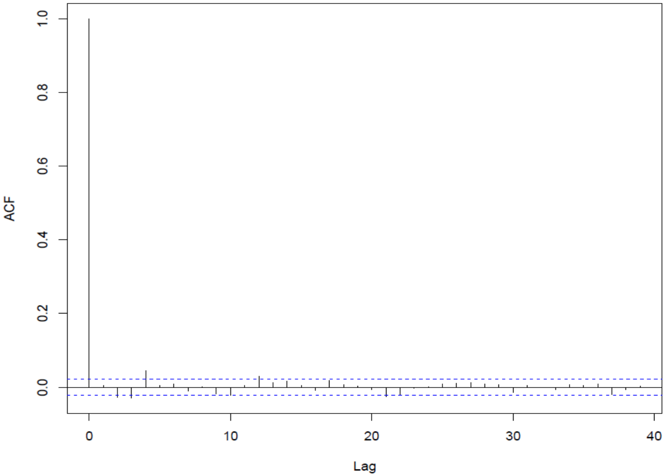
$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

- The correlation of time t's return and time $(t + j)$'s return:

$$\rho_{R,j} = \frac{Cov(R_t, R_{t+j})}{\sqrt{Var(R_t) \cdot Var(R_{t+j})}}$$

- The Auto-Correlation Function of return: $\rho_{R,j}$ as a function of $j = 0, 1, 2, \ldots$

Auto-Correlation Function of Apple Returns



- There is no strong evidence showing that the simple return R_t of the Apple shares is serially correlated
- That is R_t and R_{t+j} $(j \neq 0)$ are uncorrelated
- Such a property is sometimes called linear independence
- But this property does not imply that R_t and R_{t+j} $(j \neq 0)$ are independent

- Covariance between two random variables X and Y: $Cov(X, Y)$
- The correlation between X and Y:

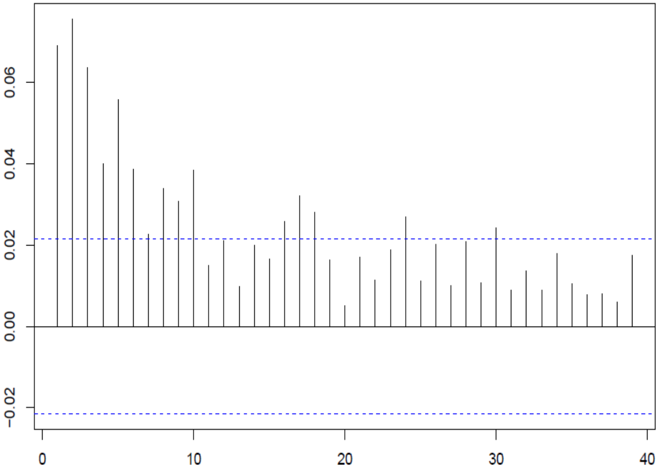
$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

- The correlation of the squared returns at time t and time $(t + j)$:

$$\rho_{R^2,j} = \frac{cov(R_t^2, R_{t+j}^2)}{\sqrt{Var(R_t^2) \cdot Var(R_{t+j}^2)}}$$

- The Auto-Correlation Function of return: $\rho_{R^2,j}$ as a function of $j = 0, 1, 2, \ldots$

Auto-Correlation Function of Squared Apple Returns (R_t^2)



- Covariance between two random variables X and Y: $Cov(X, Y)$
- The correlation between X and Y:

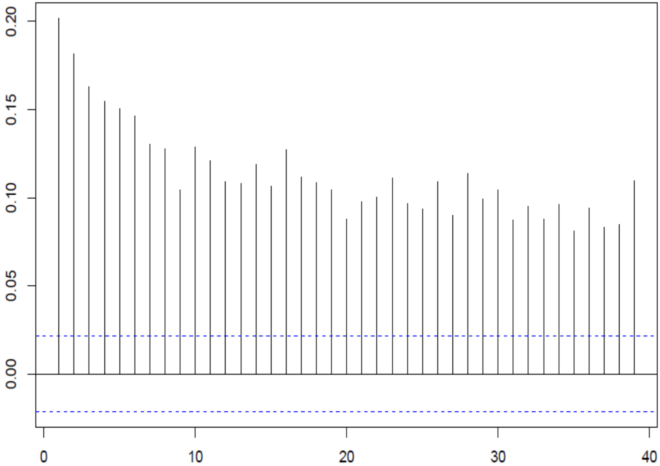
$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

- The correlation of the absolute returns at time t and time $(t + j)$:

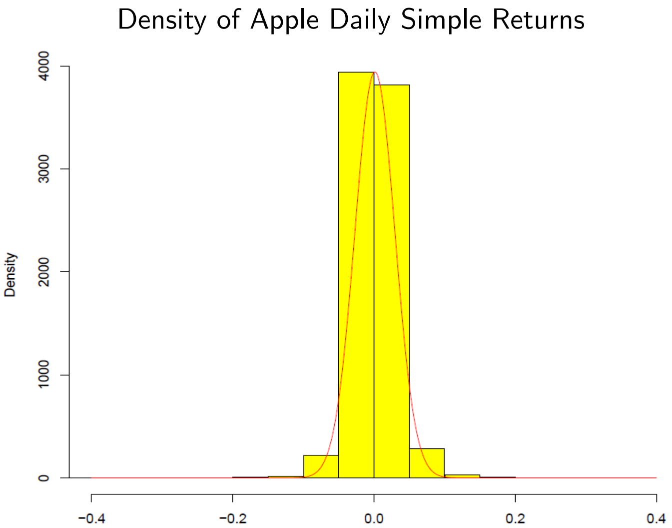
$$\rho_{R^2,j} = \frac{cov(|R_t|, |R_{t+j}|)}{\sqrt{Var(|R_t|) \cdot Var(|R_{t+j}|)}}$$

- The Auto-Correlation Function of return: $\rho_{|R|,j}$ as a function of $j = 0, 1, 2, \ldots$

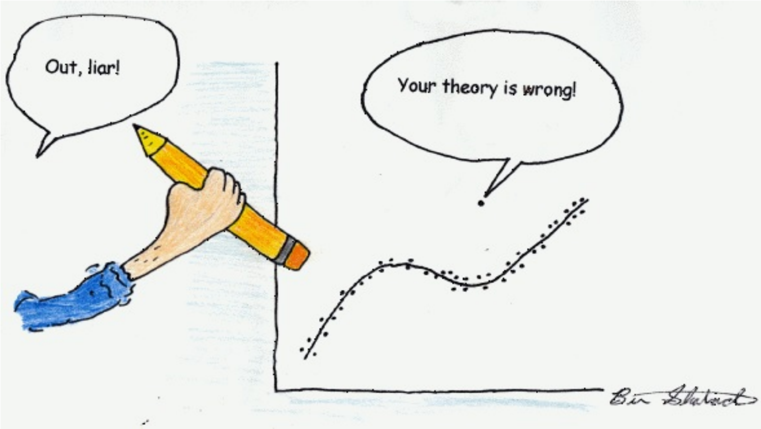
Auto-Correlation Function of Absolute Apple Returns ($|R_t|$)



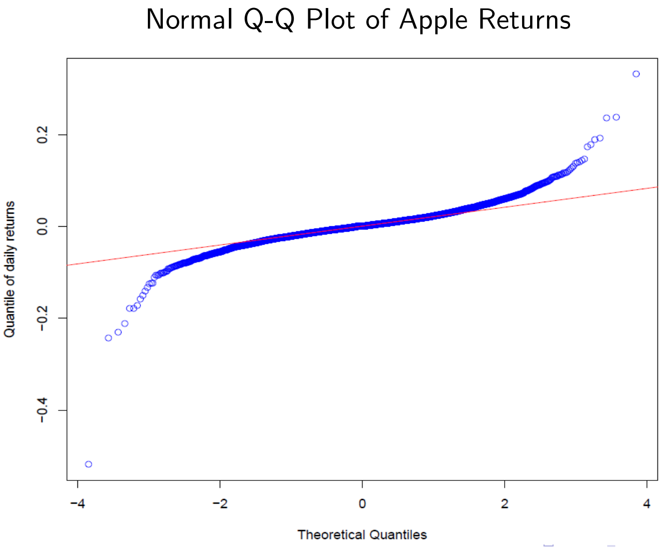
- There are strong evidence showing that the simple return R_t of the Apple shares is not independent
- That is R_t and R_{t+j} $(j \neq 0)$ are dependent
- Because nonlinear function(s) of R_t and R_{t+j} are correlated
- Such a property is sometimes called non-linear dependence


Density of Apple Daily Simple Returns
Density
0
1000
2000
3000
4000
−0.4
−0.2
0.0
0.2
0.4

- The financial returns usually have heavier tails than the normal random variable
- The main reason is because the "outliers" show up in the financial data more frequently than the normal distribution predicts
- Moreover, asymmetry often shows up in the density of the financial returns
- That is another difference between the financial returns and the normal random variable
- Therefore, the simple normal/Gaussian distribution is inappropriate to model the financial returns

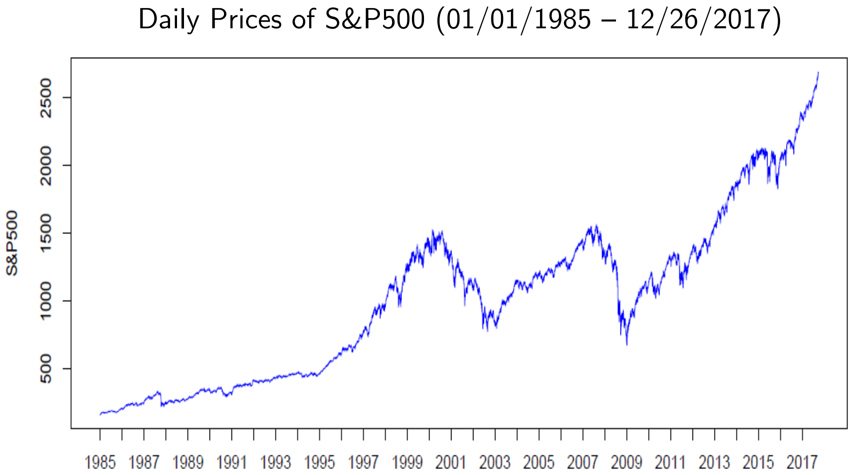

Out, liar!
Your theory is wrong!

- Random variable X has CDF $F_X(\cdot)$
- Random variable Y has CDF $F_Y(\cdot)$
- If X and Y have the same distribution, then $F_X(x) = F_Y(x)$ for any $x \in R$
- If X and Y have the same distribution, then $F_X^{-1}(u) = F_Y^{-1}(u)$ for any $u \in (0, 1)$
- Let $X = R_t$ and Y be a normal random variable
- Comparing $F_X^{-1}(u)$ with $F_Y^{-1}(u)$ for different values of $u \in (0, 1)$ gives us information whether R_t is normally distributed

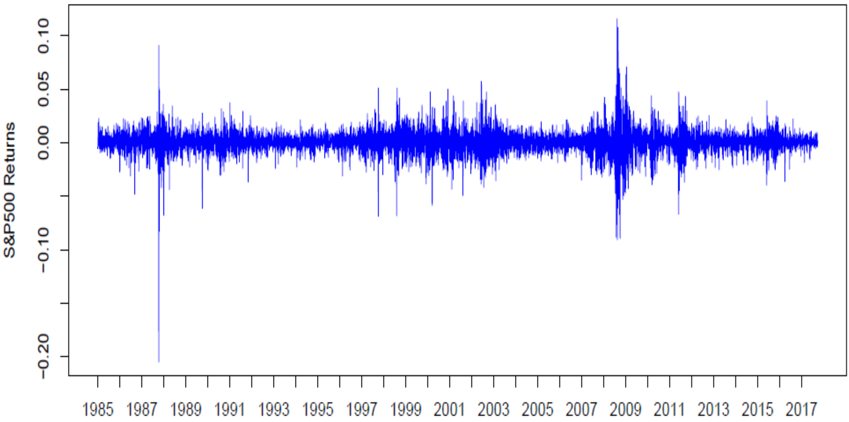

Normal Q-Q Plot of Apple Returns
Quantile of daily returns
0.2
0.0
−0.2
−0.4
−4
−2
0
2
4
Theoretical Quantiles

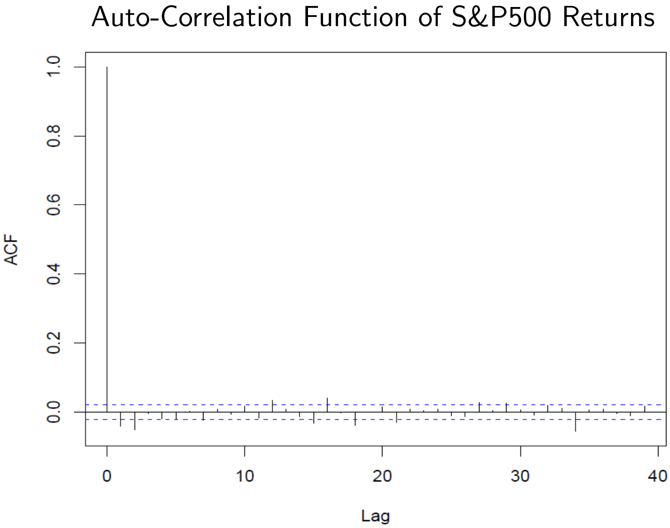
The S&P500

- We next investigate the prices of the S&P500
- The S&P500 is a value-weighted index of the prices of the 500 large-cap common stocks actively traded in the United States
- Its present form has been published since 1957, but its history dates back to 1923 when it was a value-weighted index based on 90 stocks
- It is regarded as a bellwether for the American economy
- Many mutual funds, exchange-traded funds, pension funds etc. are designed to track the performance of S&P500

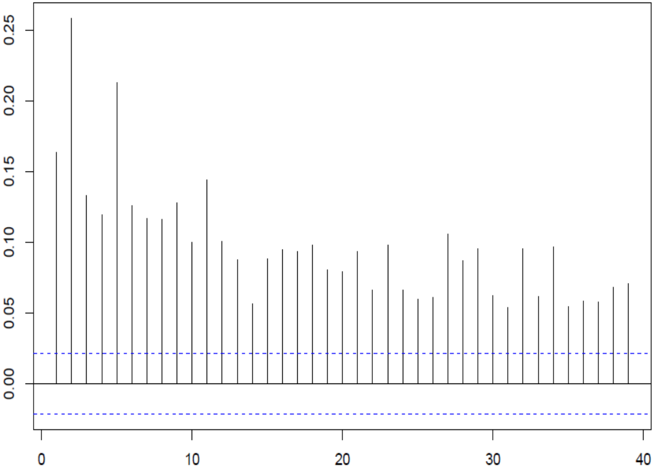

Daily Prices of S&P500 (01/01/1985 – 12/26/2017)
S&P500
500
1000
1500
2000
2500
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015
2017

Daily Returns of S&P500 (01/01/1985 – 12/26/2017)

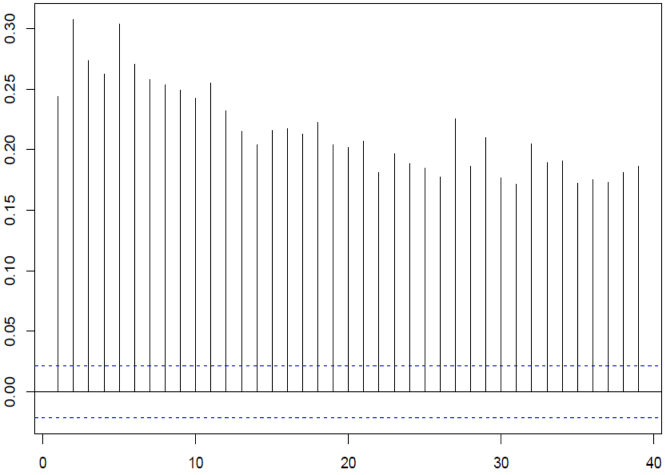


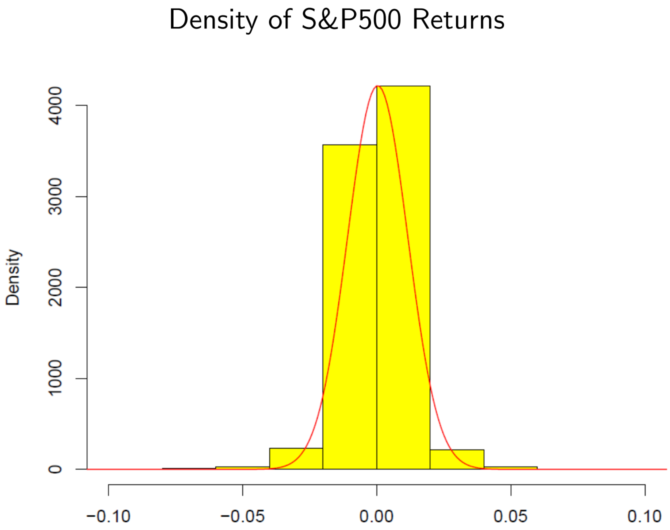

Auto-Correlation Function of S&P500 Returns
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
Lag

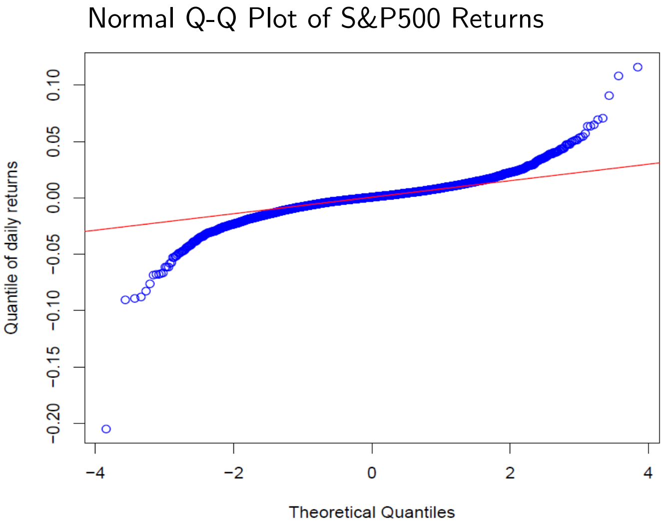
Auto-Correlation Function of Squared S&P500 Returns (R_t^2)



Auto-Correlation Function of Absolute S&P500 Returns ($|R_t|$)




Density of S&P500 Returns
Density
0
1000
2000
3000
4000
−0.10
−0.05
0.00
0.05
0.10


Normal Q-Q Plot of S&P500 Returns
Quantile of daily returns
Theoretical Quantiles
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−4
−2
0
2
4

Stylized Facts

- Trend in prices
- Stationarity
- (Non-linear) Dependence
- Heavy tails
- Asymmetry
- Volatility clustering

Course Description

- It is almost impossible to draw any useful conclusion on the joint distribution of a time series without further assumption on the underlying data generating process
- Stationarity is a useful assumption when we are interested in the mean and auto-covariance of the time series
- Another approach is to assume that the financial data is generated from some statistical models
- Financial risk analysis can be conveniently conducted using these models
- The statistical models in finance are motivated to capture the key features of the data

Course Description

- The fourth topic is the **Heteroscedastic Volatility Models**
- The fifth topic is the **Portfolio Theory**
- The sixth topic is the **Factor Pricing Model**

Course Description - prerequisite

- Required prerequisite - probability and statistics
- This is not the right course if you have never learned any of prob./stat. concepts such as, for example,
  - random variable,
  - probability density function (pdf) and distribution function (CDF),
  - expectation,
  - statistical independence.
- Course materials are uploaded in the course webpage
  - maybe updated

Course Motivation

- Why financial econometrics?
  - to confirm economics/financial theory
  - to develop a new theory
  - to manage risk
  - to make money?
- What do we need?
  - financial market data (and statistical software)
  - probability and statistics
  - time series concepts
  - econometric models and methods
  - financial theory

Statistical software - R program

- This course does not make (and require) you to be an R programmer
  - The main idea of modern statistical computing is "sharing" codes and "building upon" existing codes.
  - For each R exercises, a benchmark R coding will be provided. You will be able to "replicate" all the results in most cases.
  - It will help you to visualize and more clearly understand course materials - "replication" exercises.
  - Some guidance will be provided oftentimes, but currently there is no plan for independent R sessions.
- Available Resources from UCLA
  - https://stats.idre.ucla.edu/r/

Tentative Course Schedule

- Course Introduction, Understanding Financial Returns
- Review of Concepts in Probability and Statistics
- Time Series Concepts, Descriptive Statistics
- **Mid-term Exam**
- Conditional Volatility Models
- Portfolio Theory
- Factor Pricing Model and CAPM
- Some Additional Topics and Review