Financial econometrics

Chapter 1, Introduction

Overview

Section 1, Introduction

What is econometrics?

- What is econometrics?

  Literal meaning is ‘measurement in economics’.

  “Econometrics may be defined as the quantitative analysis of actual economic phenomena based on the concurrent development of **theory** and **observation**, related by appropriate methods of **inference**”.

[1] Paul A. Samuelson, Tjalling C. Koopmans and J. Richard N. Stone. ‘Report of the evaluative committee for Econometrica’. In: *Econometrica* 22.2 (1954), pp. 141–146

What is econometrics? (cont'd)

"A word of explanation regarding the term econometrics may be in order. Its definition is implied in the statement of the scope of the [Econometric] Society, in Section I of the Constitution, which reads: 'The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics...' [...].
Its main object shall be to promote studies that aim at a **unification of the theoretical-quantitative and the empirical-quantitative approach to economic problems** and that are penetrated by constructive and rigorous thinking similar to that which has come to dominate in the natural sciences. [...]
But there are several aspects of the quantitative approach to economics, and no single one of these aspects, taken by itself, should be confounded with econometrics. Thus, econometrics is by no means the same as economic statistics. Nor is it identical with what we call general economic theory, although a considerable portion of this theory has a definitely quantitative character. Nor should econometrics be taken as synonymous with the application of mathematics to economics. Experience has shown that each of these **three view-points**, that of **statistics, economic theory, and mathematics**, is a necessary, but not by itself a sufficient, condition for a real understanding of the quantitative relations in modern economic life. It is the **unification of all three** that is powerful. And it is **this unification that constitutes econometrics**."

[1] Ragnar Frisch. 'Editor's Note'. In: *Econometrica* 1.1 (1933), pp. 1–4

What is financial econometrics?

- What is econometrics?

  ...not an easy question!

  The application of statistical and mathematical techniques to problems in finance (Brooks, 2014).

  "...Broadly speaking, financial econometrics is to study quantitative problems arising from finance. It uses statistical techniques and economic theory to address a variety of problems from finance" (Fan, 2004).

Definition (Financial econometrics)

Financial econometrics is the application of statistical methods to financial market data. [...] It differs from other forms of econometrics because the emphasis is usually on analyzing the prices of financial assets traded at competitive, liquid markets (Wikipedia).

What is financial econometrics? (cont'd)

Objectives of financial econometrics:

- Combining finance theory with statistical theory.
- Transferring a theoretical (financial) knowledge into a (financial) econometric model.
- Modelling financial data.
- Predicting financial variables as well as relations thereof.
- Developing specific model well-adapted to the stylized facts of financial data. Example: GARCH models.

What is financial econometrics? (cont'd)

It is useful?

> "People working in the finance industry or research in the finance sector often use **econometric techniques** in a range of activities - for example, in support of **portfolio management** and in the **valuation of securities**. Financial econometrics is essential for risk management when it is important to know how often 'bad' investment outcomes are expected to occur over future days, weeks, months and years" (Wikipedia).

Examples of the kind of problems that may be solved by an econometrician

1. Testing whether financial markets are weak-form informationally efficient.
2. Testing whether the CAPM or APT represent superior models for the determination of returns on risky assets.
3. Measuring and forecasting the volatility of bond returns.
4. Explaining the determinants of bond credit ratings used by the ratings agencies.
5. Modelling long-term relationships between prices and exchange rates.
6. Determining the optimal hedge ratio for a spot position in oil.

Examples of the kind of problems that may be solved by an econometrician (cont'd)

7. Testing technical trading rules to determine which makes the most money.
8. Testing the hypothesis that earnings or dividend announcements have no effect on stock prices.
9. Testing whether spot or futures markets react more rapidly to news.
10. Forecasting the correlation between the returns to the stock indices of two countries.

Remark

There is sometimes a confusion between the following terms

- **Financial econometrics** bridges the gap between financial economics, statistics and mathematical finance, e.g. *Econometrica*.
- **Empirical finance** covers all the finance studies based on data analysis, e.g. *Journal of Empirical Finance*.
- **Financial economics** is a "highly empirical discipline, perhaps the most empirical among the branches of economics and even among the social sciences in general" (Campbell, Lo and MacKinlay, 1996), e.g. *Journal of Financial Economics*.

Elements of financial econometrics

(Financial) econometrics is fundamentally based on four elements:

- A **question** to address, of financial/economic relevance
- A **sample** of data
- An econometric **model**
- An **estimation** method
- Some **inference** methods

What are the special characteristics of financial data?

- *Frequency & quantity of data*

  Stock market prices are measured every time there is a trade (e.g. transaction price), a quote posted (e.g. bid/ask price), or somebody posts -or cancels- a market order (e.g. mid-price).

- *Quality*

  Recorded asset prices are usually those at which the transaction took place. No possibility for measurement error but financial data are "noisy".

- *Market microstructure* (for high-frequency data)

  "A variety of frictions inherent in the trading process: bid-ask bounces, discreteness of price changes, differences in the informational content of price changes" [...] (e.g. Aït-Sahalia and Yu, 2008)

Sample

Question: Why using a sample?

- Let us assume that we want to study a **characteristic / property** x of the **individuals** of a **population**.
- The individuals (units) of the population are not necessarily some persons: it can be firms, assets, countries, time index etc...
- The characteristic x may be **quantitative** (price, returns, total asset, etc.) or **qualitative** (default, sector, etc.).
- The characteristic x may be **stochastic** or **deterministic**.

Sample (cont'd)

Definition (Population)

A population can be defined as including all people or items with the characteristic one wishes to understand.

1. In most of cases, it is impossible to observe the entire statistical population, due to cost constraints, time constraints, constraints of geographical accessibility.
2. A researcher would instead observe a statistical **sample from the population** in order to attempt to **learn something about the population as a whole**.

Sample (cont'd)

In most of cases, the sample is randomly selected:

Definition (Probability sampling)

A probability sampling is a **sampling method** in which every unit in the population has a chance (greater than zero) of being selected in the sample.

Consequence: a sample is a collection of random variables even if the characteristic x is deterministic.

$$\text{sample: } \{X_1, X_2, ...X_N\}$$

Sample (cont'd)

In most of cases, the sample is random:

Random sample (dataset)

The result of the probability sampling is a **random sample**, i.e. a collection of random variables $X_1, X_2, ...X_N$. In general, only **one realization** of the sample is available: this is your data set.

$$\{x_1, x_2, ...x_N\}$$

Sample (cont'd)

Sample (cont'd)

Remark: when ones consider two or more variables, the notion of population is replaced by the concept of Data Generating Process (DGP).

Definition (Data generating process)

A Data Generating Process (DGP) is the joint probability distribution that is supposed to characterize the entire population from which the data set has been drawn.

Example

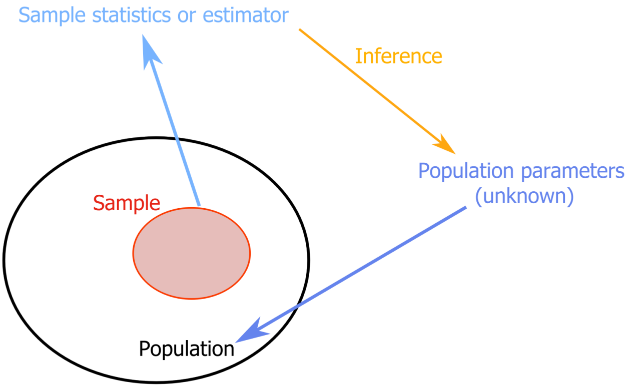
Let us assume that there is a linear relationship between the returns R_t and $R_{m,t}$ in the population, such that

$$\mathbb{E}\left(R_t | R_{m,t} = r_m\right) = \alpha + \beta r_m$$

This relationship is the Data Generating Process for R_t.

Sample (cont'd)

The **challenge** of econometrics is to draw conclusions about a population (or DGP) after observing only **one** realization $\{x_1, x_2, ..., x_N\}$ of a random sample (your data set).



Types of data and notation

- There are 3 types of data that econometricians might use for analysis:

  1. Time series data
  2. Cross-sectional data
  3. Panel data, a combination of 1. & 2.

- The data may be quantitative (e.g. exchange rates, stock prices, number of shares outstanding), or qualitative (e.g. day of the week).
- The data may be continuous or discrete.
- The data may be cardinal, ordinal, or nominal.

Types of data and notation (cont'd)

Time series data

- Data for a single entity (asset, firm, etc.) collected at *multiple* time periods, i.e. repeated observations of the same variables (price, volume, etc.).
- Order of data is important.
- Observations are typically not independent over time.

Examples of problems that could be tackled using time series data

– How the value of a country's stock index has varied with that country's macroeconomic fundamentals.

– How the value of a company's stock price has varied when it announced the value of its dividend payment.

Types of data and notation (cont'd)

Cross-sectional data

- Data on one or more variables collected at a single point in time,
- Data for different entities: assets, portfolios, firms, and so forth.
- No time dimension (even if the date of data collection varies somewhat across units, it is ignored)
- Order of data does not matter

Examples of problems that could be tackled using a cross-sectional data:

- The relationship between company size and the return to investing in its shares
- The relationship between a country's GDP level and the probability that the government will default on its sovereign debt.

Types of data and notation (cont'd)

Panel (longitudinal) data

- Data for multiple entities (asset, firm, etc.) in which outcomes and characteristics of each entity are observed at multiple points in time.
- Combine cross-sectional and time series issues.
- Present several advantages with respect to cross-sectional and time series data (depending on the question of interest!).

It is common to denote each observation by the letter t and the total number of observations by T for time series data, and to denote each observation by the letter i and the total number of observations by N for cross-sectional data.

$$X_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T$$

Continuous and discrete data

- Continuous data can take on any value and are not confined to taking specific numbers.
- Their values are limited only by precision.
  - For example, the rental yield on a property could be 6.2%, 6.24%, or 6.238%.
- On the other hand, discrete data can only take on certain values, which are usually integers
  - For instance, the number of people in a particular underground carriage or the number of shares traded during a day.
- They do not necessarily have to be integers (whole numbers) though and are often defined to be count numbers.
  - For example, until recently when they became 'decimalised', many financial asset prices were quoted to the nearest 1/16 or 1/32 of a dollar.

Cardinal, ordinal and nominal numbers

- Another way in which we could classify numbers is according to whether they are cardinal, ordinal, or nominal.
- *Cardinal numbers* are those where the actual numerical values that a particular variable takes have meaning, and where there is an equal distance between the numerical values.
  - Examples of cardinal numbers would be the price of a share or of a building, and the number of houses in a street.
- *Ordinal numbers* can only be interpreted as providing a position or an ordering.
  - Thus, for cardinal numbers, a figure of 12 implies a measure that is 'twice as good' as a figure of 6. On the other hand, for an ordinal scale, a figure of 12 may be viewed as 'better' than a figure of 6, but could not be considered twice as good. Examples of ordinal numbers would be the position of a runner in a race.

Cardinal, ordinal and nominal numbers (cont'd)

- Nominal numbers occur where there is no natural ordering of the values at all.
  - Such data often arise when numerical values are arbitrarily assigned, such as telephone numbers or when codings are assigned to qualitative data (e.g. when describing the exchange that a US stock is traded on).
- Cardinal, ordinal and nominal variables may require different modelling approaches or at least different treatments, as should become evident in the subsequent chapters.

Econometric model

Definition (Econometric model)

A **model** specifies the statistical relationship that is believed to hold between the various economic quantities pertaining to a particular economic phenomenon under study.

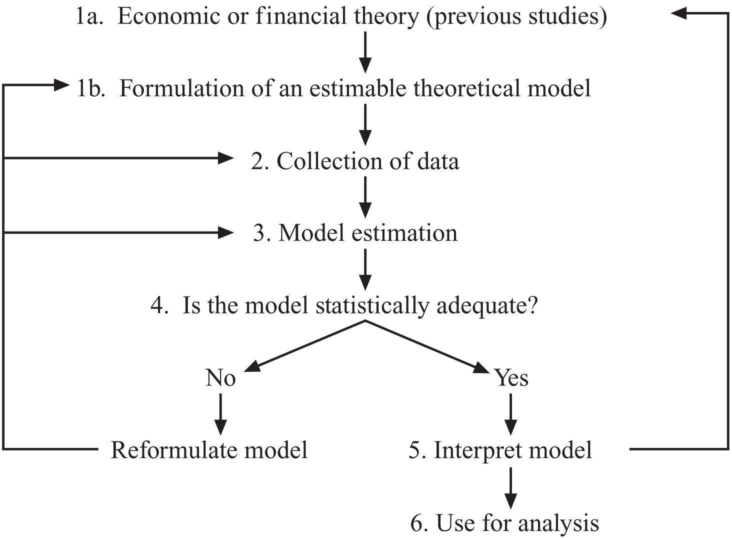
We can distinguish:

1. **Parametric model:** the relationship (joint probability distribution) between the dependent variable(s) Y and the explicative variables X is fully characterized by a set of parameters θ

$$Y = f(X; \theta) + \varepsilon$$

   where link function $f(\cdot)$ is assumed to be known.
2. **Non parametric and semi-parametric models:** the link function can not be described using a finite number of parameters. The link function is assumed to be unknown and has to be estimated.

Steps in the formulation of econometric models



Econometric approach

The general approach of (financial) econometrics is the following:

1. Specification of the model (Chapters 2, 3)
2. Estimation of the model (Chapters 2, 3)
3. Diagnostic tests (Chapter 4)
   - Significance test
   - Specification tests
   - Backtesting
   - etc.
4. Interpretation and use of the model, e.g. forecasting and answering research questions (Chapter 5).

Section 2, Prices and returns

Liquidity

Illiquid markets
A market in which that is difficult to sell assets because of their expense, lack of interested buyers, or some other reason (e.g. highly regulated).

Liquidity (cont'd)

Liquid markets

Quantitative financial research mainly focuses on 'liquid' financial markets, i.e. organized markets where transactions are frequent and the number of actors is large.

Example ('Liquid' markets)

Some example of 'liquid' markets: foreign exchange market, organized futures markets, stock index markets and the market for large stocks

- On these markets, prices are recorded several times a minutes (up to nanosecond precision) or after each event (tick-by-tick data).
- A liquid market has plenty of buyers and sellers. In such a market, it's easy to get buy/sell orders filled at a reasonable price.
- The typical trading mechanism is the electronically-driven Limit Order Book (LOB).

Price

Definition (Price)

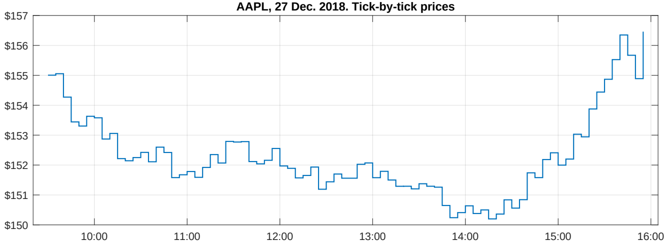
Let P_t denotes the price of an asset at time t.

The prices may be observed and recorded at:

1. An irregular (sampling) frequency: tick-by-tick price observations, volume-event observations (prices observed when the volume exceeds a given threshold), price-events (transactions associated with significant price changes), etc.
2. A regular (sampling) frequency: the prices are observed every m periods of time.

Price (cont'd)


AAPL, 27 Dec. 2018. 5-min prices
$157
$156
$155
$154
$153
$152
$151
$150
10:00
11:00
12:00
13:00
14:00
15:00
16:00


AAPL, 27 Dec. 2018. Tick-by-tick prices
$157
$156
$155
$154
$153
$152
$151
$150
10:00
11:00
12:00
13:00
14:00
15:00
16:00

Prices (cont'd)

High-frequency data	**Low frequency data**
tick-by-tick	**daily**
...	weekly
10 sec	**monthly**
1 min	quarterly
5 min	sami-annual
10 min	annual
30 min	**...**
1 hour	
...	
less than a day	

Price, example

Example (Sampling frequency for prices)

Plot the time-series of (closing) prices for Microsoft Corporation (Ticker: MSFT) up to June 30th, 2021, sampled at daily and weekly frequency respectively.

Note: The data are available at `finance.yahoo.com`

- Ultra-high-frequency and Tick-by-tick data is of huge size, requires proper hardware to be recorded and thus sold at expensive rates. However for certain markets (e.g. cryptocurrencies) high-frequency data (e.g. up to a minute) is fairly easy to retrieve.

Price, example (cont'd)

Figure: Daily prices

Price, example (cont'd)

Figure: Weekly prices

Prices

For a given sampling frequency (e.g., daily), the prices can be recorded at different times within the observational period.

For instance, for a daily sampling frequency, we can consider:

- The **opening price** is the price at which a security first trades upon the opening of an exchange on a trading day.
- The **closing price** is the final price at which a security is traded on a given trading day, representing the most up-to-date valuation of a security
- The **highest/lowest** price of the trading day.
- The **adjusted closing prices** have been treated to correct for splits and dividends.

In general, we consider the **end-of-period** (closing) price, for a given sampling frequency.

Prices (cont'd)

Time Period: Jan 01, 2021 - Sep 12, 2021 ⌄ Show: Historical Prices ⌄ Frequency: Daily ⌄ Apply

Currency in USD Download

Date	Open	High	Low	Close*	Adj Close**	Volume
Sep 10, 2021	298.42	299.92	295.38	295.71	295.71	19,619,400
Sep 09, 2021	300.82	302.14	297.00	297.25	297.25	19,927,000
Sep 08, 2021	299.78	300.61	297.47	300.21	300.21	15,046,800
Sep 07, 2021	301.01	301.09	298.20	300.18	300.18	17,180,400
Sep 03, 2021	300.99	302.60	300.26	301.14	301.14	14,747,900
Sep 02, 2021	302.20	303.36	300.18	301.15	301.15	16,285,600
Sep 01, 2021	302.87	305.19	301.49	301.83	301.83	18,983,800
Aug 31, 2021	304.42	304.50	301.50	301.88	301.88	26,285,300
Aug 30, 2021	301.12	304.22	301.06	303.59	303.59	16,348,100
Aug 27, 2021	298.99	300.87	296.83	299.72	299.72	22,597,000
Aug 26, 2021	300.99	302.43	298.95	299.09	299.09	17,666,100

Figure: MSFT, OHLC prices.

Prices (cont'd)

Time Period: Jan 01, 2021 - Sep 12, 2021 Show: Historical Prices Frequency: Daily Apply

Currency in USD Download

Date	Open	High	Low	Close*	Adj Close**	Volume
Sep 10, 2021	298.42	299.92	295.38	295.71	295.71	19,619,400
Sep 09, 2021	300.82	302.14	297.00	297.25	297.25	19,927,000
Sep 08, 2021	299.78	300.61	297.47	300.21	300.21	15,046,800
Sep 07, 2021	301.01	301.09	298.20	300.18	300.18	17,180,400
Sep 03, 2021	300.99	302.60	300.26	301.14	301.14	14,747,900
Sep 02, 2021	302.20	303.36	300.18	301.15	301.15	16,285,600
Sep 01, 2021	302.87	305.19	301.49	301.83	301.83	18,983,800
Aug 31, 2021	304.42	304.50	301.50	301.88	301.88	26,285,300
Aug 30, 2021	301.12	304.22	301.06	303.59	303.59	16,348,100
Aug 27, 2021	298.99	300.87	296.83	299.72	299.72	22,597,000
Aug 26, 2021	300.99	302.43	298.95	299.09	299.09	17,666,100

Figure: MSFT, OHLC prices.

Prices and returns

Fact (prices vs returns)

Although **prices** are what we observe in financial markets, most empirical studies are based on **returns**.

- Returns are a complete and scale-free summary of investment opportunity
- Returns are easier to handle in practice
- Returns have more attractive statistical proprieties (are, in general, **stationary**)

[1] John Y. Campbell, Andrew W. Lo and A. Craig MacKinlay. *The econometrics of financial markets*. Princeton University press, 1996

Prices and returns (cont'd)

There are several definitions of an asset return, we are mainly interested in the following:

1. One-period simple gross return
2. One-period simple net return
3. Multi-period simple return
4. Continuously compounded return (log-return)
5. Portfolio return
6. Excess return

Gross and net retun

Definition (One-period simple gross return)

Holding an asset for one period from date $t-1$ to date t results in a simple **gross return**:

$$1 + R_t = \frac{P_t}{P_{t-1}} \qquad \text{or} \qquad P_t = P_{t-1}\,(1 + R_t)$$

Definition (One-period simple net return)

Holding an asset for one period from date $t-1$ to date t results in a simple **net return**:

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

Multi-period return

Definition (Multi-period simple return)

Holding the asset for k periods between dates $t-k$ and t gives a k-period simple gross return:

$$1 + R_t(k) = \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \cdots \times \frac{P_{t-k+1}}{P_{t-k}}$$
$$= (1+R_t)(1+R_{t-1}) \times \cdots \times (1+R_{t-k+1})$$
$$= \prod_{j=0}^{k-1} (1+R_{t-j})$$

- This is called compounded return.
- The k-period simple net return is $R_t(k) = \frac{P_t - P_{t-k}}{P_{t-k}}$.

Annualization

- 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.

Definition (Annualized return)

If the asset is held for k years, the annualized (average) return is defined as:

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

- This is a geometric mean of the k one-period simple gross returns involved.

Example

Example

Assume that the interest rate of a bank deposit is 10% per annum and the initial deposit is \$1.00. If the bank pays interest once a year, then the net value of the deposit becomes $\$1(1+0.1) = \$1.1$ one year later. If the bank pays interest semiannually, the 6-month interest rate is $10\%/2 = 5\%$ and the net value is $\$1(1+0.1/2)^2 = \$1.1025$ after the first year. In general, if the bank pays interest m times a year, then the interest rate for each payment is $10\%/m$ and the net value of the deposit becomes $\$1(1+0.1/m)^m$ one year later.
In particular for $m \to \infty$, the net value approaches \$1.1052, which is $\exp(0.1)$ and referred to as the result of **continuous compounding**.

Continuous compounding

In general, the net asset value A of continuous compounding is

$$A = C \exp\left(r \times n\right),$$

where r is the interest rate per annum, c is the initial capital, and n is the number of periods. Analogously,

$$C = A \exp\left(-r \times n\right).$$

Type	Number of payments	Interest rate per period	Net value
Annual	1	0.100	1.10000
Semiannual	2	0.050	1.10250
Quarterly	4	0.025	1.10381
Monthly	12	0.008	1.10471
Weekly	52	$1/52$	1.10506
Daily	365	$1/365$	1.10516
Continuously	∞		1.10517

Table: Effect of compounding $1 for one year at 10% interest rate per annum.

Logreturn

Definition (Log-return)

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

$$r_t = \log P_t - \log P_{t-1} = \log \frac{P_t}{P_{t-1}} = \log\left(1 + R_t\right),$$

- The continuously compounded multi-period return is simply the sum of the continuously compounded one-period returns involved in the period:

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

Logreturn (cont'd)

- The definition of logreturn also reads:

$$R_t = e^{r_t} - 1.$$

Question: if X is normally distributed, what is the distribution of e^X?

- When net returns are small:

$$r_t = \log(1 + R_t) \approx R_t,$$

however $r_t \leq R_t$.

Logreturn (cont'd)

X	$x = \log(1+X)$	err. $= \lvert\frac{x-X}{X}\rvert$
-0.150	-0.163	7.7%
-0.125	-0.134	6.4%
-0.100	**-0.105**	**5.1%**
-0.075	-0.078	3.8%
-0.050	-0.051	2.5%
-0.025	-0.025	1.3%
0.000	**0.000**	**0.0%**
0.025	0.025	1.2%
0.050	0.049	2.5%
0.075	0.072	3.7%
0.100	**0.095**	**4.9%**
0.125	0.118	6.1%
0.150	0.140	7.3%

Table: Validity of the approximation.

- Rule of thumb: for $|R_t| \leq 10\%$ the error between $r_t = \log(1+R_t)$ and R_t is approximately within 5%.

Logreturn (cont'd)

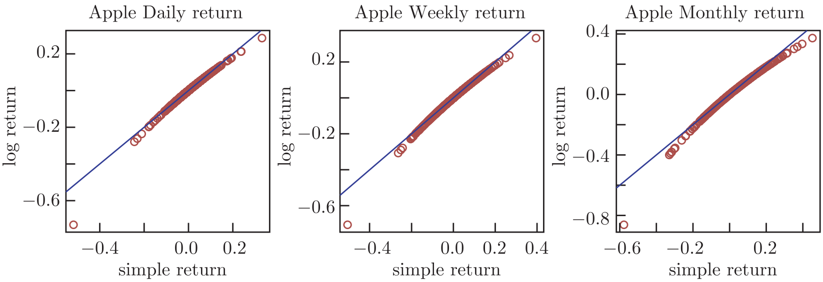


Figure: Plots of log returns against simple returns of the Apple Inc share prices in January 1985 - February 2011. The blue straight lines mark the positions where the two returns are identical (Fan and Yao, 2017).

Adjustment for dividends

- Many assets, pay dividends to their share-holders from time to time.
- A dividend is typically allocated as a fixed amount of cash per share
- adjustments must then be made in computing returns to account for the contribution towards the earnings from dividends

Returns **adjusted for dividends** are computed as follows:

$$R_t = (P_t + D_t)\,/P_{t-1} - 1,$$
$$r_t = \log(R_t + D_t) - \log P_{t-1},$$

$$R_t(k) = (P_t + D_t + D_{t-1} + \cdots + D_{t-k+1})\,/P_{t-k} - 1,$$
$$r_t(k) = r_t + \cdots + r_{t-k+1} = \sum_{j=0}^{k-1} \log\left(\frac{P_{t-j} + D_{t-j}}{P_{t-j-1}}\right),$$

with $D_{t-j} = 0$ if no dividend is paid on day $t - j$.

Example

Example

The closing price of the equity Apple Inc (AAPL) was equal to \$154.10 on September 10, 2021 and \$148.97 on September 11, then the daily return is equal to

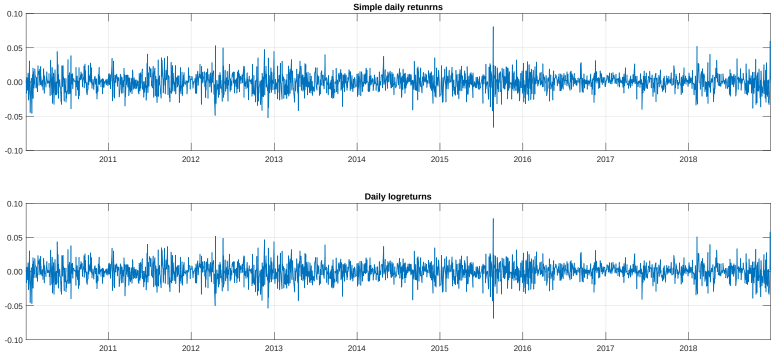
$$R_t = \frac{148.97 - 154.10}{154.10} = -0.0333 = -3.33\%$$

$$r_r = \log \frac{148.97}{154.10} = \log 148.97 - \log 154.10 = -0.0339 = -3.34\%$$

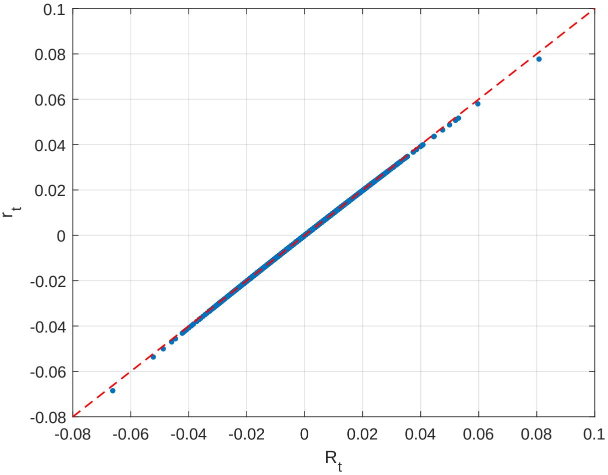
Example

Example

Consider a sample of daily closing prices for the equity Apple Inc (AAPL) from Jan. 1, 2010 to Dec. 31, 2018. Compare the (simple) daily returns and the log-(daily) returns.



Prices and returns (cont'd)



Logreturn (cont'd)

A disadvantage of using logreturns

- There is a disadvantage of using the log-returns. The simple return on a portfolio of N assets is a weighted average of the simple returns on the individual assets:

$$R_p(t) = \sum_{i=1}^{N} w_i R_i(t).$$

- But this does not work with the logreturns:

$$\begin{aligned} r_p(t) &= \log(1 + R_p(t)) \\ &= \log\left(1 + \sum_{i=1}^{N} w_i R_i(t)\right) \neq \sum_{i=1}^{N} w_i \log(1 + R_i(t)) = \sum_{i=1}^{N} w_i r_i(t). \end{aligned}$$

Excess return

Definition (Excess return)

In many applications, it is convenient to use an excess return defined as

$$\text{excess return} = R - R_m,$$

where R_m is a generic reference rate.

Commonly used reference rates:

- LIBOR (London Interbank Offered Rate): the average interest rate that leading banks in London charge when lending to other banks.
- Market portfolio: S&P 500 index or CRSP value-weighted index
- Yield excess: excess yield defined as the difference between the yield of a bond and the yield of a reference bond such as a US treasury bill with a similar maturity.

Real vs. nominal series

- The general level of prices has a tendency to rise most of the time because of inflation.
- We may wish to transform nominal series into real ones to adjust them for inflation.
- This is called *deflating* a series or displaying a series *at constant prices*.
- We do this by taking the nominal series and dividing it by a price deflator (assuming that the base figure is 100):

  $\text{real series}_t = \text{nominal series}_t \times 100/\text{deflator}_t$

- We only deflate series that are in nominal price terms, not e.g. quantity terms.

Section 3,
Limit-order book

Limit-order book

Question: where do prices come from?

- In the following, we will see the **trading system** that determines and drives prices.
- This is also the source that generates the data!

The old days: floor broking

Figure: The NYSE floor in 1941, the day after Pearl Harbor.

Nowadays: electronic trading

Figure: The NYSE floor on March 20th, 2020.

Limit-order book

- Trading system used by most exchanges globally:
  - E.g. NYSE, NASDAQ, TSE, HKEX, LSE.
- Trading mechanism is rather simple.
- All the investors can participate.
- Largest volume outside any other source:
  - Trades and quotes
  - Up to microsecond granularity
  - Multiple levels (bid-ask, Level II and full-depth data)
  - Updates on submitted orders

Example (an average trading day)

Average day for Volvo B, between June 2010 and May 2013:
15.850 submitted orders, of which 2.406 traded and 13.446 cancelled.

Limit-order book (cont.)

LOBs are called ‘order-driven’ markets: market participants post two types of orders.

Definition (Limit-order)

A **limit** order is an order to trade a certain amount of a security at a given price.

- Limit orders are posted to an electronic trading system and the state of outstanding limit orders can be summarized by stating the quantities posted at each price level: this is known as the **limit order book**.
- The lowest price for which there is an outstanding limit sell order is called the **ask price** and the highest buy price is called the **bid price**.

Limit-order book (cont.)

Definition (Market order)

A **market** order is an order to buy/sell a certain quantity of a security at the best available price.

- A market order is an order to buy/sell (a certain quantity of) the asset at the best available price in the limit order book.
- When a market order arrives it is matched with the best available price in the limit order book and a **trade** occurs.

Limit-order book (cont.)

- A limit order sits in the order book until it is either executed against a market order or it is **canceled**. A limit order can be canceled at any time.
- Buy limit orders are matched with ask market orders. Ask limit orders are matched against bid market orders.
- A limit order may be executed very quickly if it corresponds to a price near the bid and the ask but may take a long time if the market price moves away from the requested price or if the requested price is too far from the bid/ask.
- Every order comes with its **size** (i.e. quantity of shares, goods).

A mathematical model for Limit-order book

- Limit orders can be placed on a price grid $\{1, \dots, n\}$ representing multiples of a price tick (e.g. \$0.01).
- Track the state of the OB with a continuous-time process $X(t) = (X_1(t), \cdots, X_p(t))$
- $|X_p(t)|$ is the number of outstanding limit orders at price p, $1 \le p \le n$.
- if $X_p(t) < 0$ there are $-X_p(t)$ bid orders at price p; if $X_p(t) > 0$ there are $X_p(t)$ ask orders at price p

Limit-order book (cont.)

- The **ask price** $p_A(t)$ at time t is then defined as

$$p_A(t) = \min\left(n+1, \inf\left\{X_p(t) > 0,\ p = 1, \ldots, n\right\}\right).$$

- The **bid price** $p_B(t)$ at time t is then defined as

$$p_B(t) = \max\left(0, \left\{\sup X_p(t) < 0\, p = 1, \ldots, n\right\}\right).$$

- The **mid-price** $p_M(t)$ is defined as

$$p_M(t) = \frac{p_B(t) + p_A(t)}{2}.$$

- The **bid-ask spread** $s(t)$ is defined as

$$s(t) = p_A(t) - p_B(t).$$

Limit-order book (cont.)

- A limit buy order at price level $p < p_A$ increases the quantity at level p
- A limit sell order at price level $p > p_B$ increases the quantity at level p
- A market buy order decreases the quantity at the ask price p_A
- A market sell order decreases the quantity at the bid price p_B
- A cancellation of an outstanding limit buy order at price level $p < p_A$ decreases the quantity at level p
- A cancellation of an outstanding limit sell order at price level $p > p_B$ decreases the quantity at level p

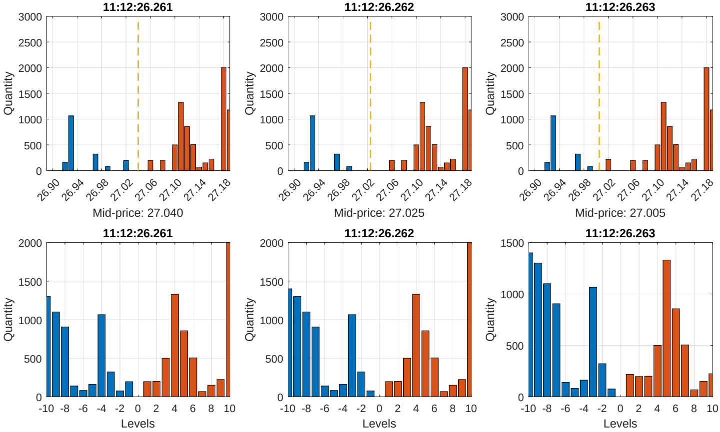
Limit-order book (cont.)

- Market efficiency and absence of arbitrages intrinsically imply that $p_B < p_A$.
  Question: How would you make a certain profit (arbitrage) if at some point you observe $p_A > p_B$?
- Moreover, keep in mind that the market design also implies the existence of a non-negative spread $s > 0$. That is, p_B is never equal to p_A (a deeper discussion is here out of scope).
- The evolution of the order book is thus driven by the incoming flow of market orders, limit orders and cancellations at different price levels.
- Incoming orders arrive more frequently in the vicinity of the current bid/ask price and the rate of arrival of these orders depends on their distance from the bid/ask price.

[1] Rama Cont, Sasha Stoikov and Rishi Talreja. 'A stochastic model for order book dynamics'. In: *Operations research* 58.3 (2010), pp. 549–563

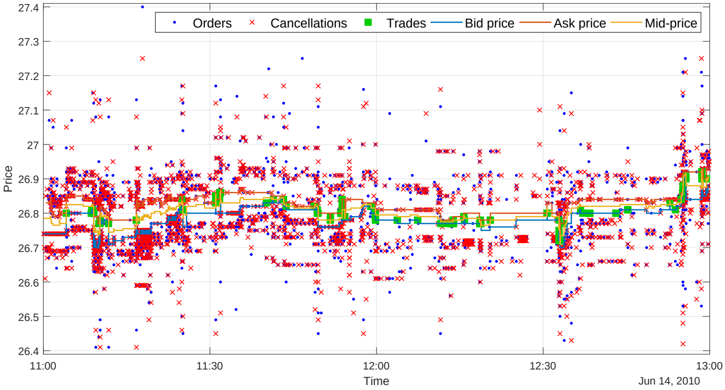
Example

Example: KESKOB, June 14^{th}, 2010.



Example (cont'd)

- A **flow perspective** reveals a true multidimensional and interacting complexity.



Example (cont.)

- A **state** perspective, i.e. limit order book.

			Bid side							
			Price				Quantity			
Time	Mid-price	Spread	Level 1	Level 2	Level 3	...	Level 1	Level 2	Level 3	...
11:12:26.261	27.040	0.04	27.02	26.99	26.97	...	196	76	322	...
11:12:26.262	27.025	0.07	26.99	26.97	26.93	...	76	322	1065	...
11:12:26.263	27.005	0.03	26.99	26.97	26.93	...	76	322	1065	...
11:12:26.558	27.005	0.03	26.99	26.97	26.93	...	76	241	1065	...
11:12:26.559	27.005	0.03	26.99	26.97	26.93	...	76	241	1065	...
11:12:26.571	27.005	0.03	26.99	26.97	26.93	...	76	241	1065	...
...	...	...	...	...	...	...	...	...	...	...

	Ask side							
	Price				Quantity			
	Level 1	Level 2	Level 3	...	Level 1	Level 2	Level 3	...
...	27.06	27.08	27.10	...	197	200	500	...
...	27.06	27.08	27.10	...	197	200	500	...
...	27.02	27.06	27.08	...	218	197	200	...
...	27.02	27.06	27.08	...	218	197	200	...
...	27.02	27.06	27.08	...	218	197	200	...
...	27.02	27.08	27.10	...	218	200	500	...
...	...	...	...	...	...	...	...	...

Example

- https://www.binance.com/en/orderbook/BTC_USDT

Section 4, Statistical properties of financial data

Notation

Definition (Distribution of returns)

We represent the (log-) returns by the real-valued random variable R_t. The distribution of R_t is called the distribution of returns, or returns' distribution

- The probability density function (pdf) of R_t is denoted $f_R(r)$.
- The cumulative density function (cdf) of R_t is denoted $F_R(r)$. such that

$$\Pr(R_t \leq r) = F_R(r) = \int_{-\infty}^{r} f_R(x)\, dx.$$

Example

Example

Let us assume the the return R_t has a normal distribution

$$R_t \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

with mean (expected return) μ and variance σ^2. Then the pdf and cdf of R_f are respectively

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

$$F_R\left(r\right) = \Phi\left(\frac{r-\mu}{\sigma}\right) = \int_{-\infty}^{r} f_R\left(x\right) dx$$

where $\phi\left(\cdot\right)$ and $\Phi\left(\cdot\right)$ respectively denote the pdf and cdf of he standard normal distribution.

Moments

Definition

The k-th **non-central moment** of R_t is given by:

$$m_k = \mathbb{E}\left(R_t^k\right) = \int_{-\infty}^{+\infty} x^k f_R(x)\, dx.$$

The k-th **central moment** of R_t is given by:

$$\mu_k = \mathbb{E}\left[\left(R_t - m_1\right)^k\right] = \int_{-\infty}^{+\infty} \left(x - m_1\right)^k f_R(x)\, dx.$$

Moments (cont'd)

In financial econometrics, the interest is mainly on the first four moments: Mean, Variance, Skewness and Kurtosis.

Security	Start	N. Obs	Mean	Std. Dev.	Skewness	Kurtosis	Minimum	Maximum
S&P 500	01/02	9845	0.023	1.062	-1.17	30.2	-22.9	10.96
IBM	01/02	9845	0.026	1.694	-0.27	12.17	-26.09	12.37
Intel	12/15	9096	0.066	2.905	-0.54	7.81	-35.06	23.41
3M	01/02	9845	0.034	1.488	-0.78	20.57	-30.08	10.92
Microsoft	03/14	5752	0.095	2.369	-0.63	14.23	-35.83	17.87
Citi-Grp	10/30	5592	0.033	2.575	0.22	33.19	-30.66	45.63

Table: Returns are in percentages and the sample period ends on December 31, 2008 (Fan and Yao, 2017).

Mean and variance

Definition (Mean and variance of returns)

The mean (expected value) of the return distribution R_t is:

$$\mu = m_1 = \mathbb{E}(R_t) = \int_{-\infty}^{+\infty} x f_R(x)\, dx.$$

The variance of the return distribution R_t is:

$$\sigma^2 = \mathbb{V}(R_t) = \mathbb{E}\left[(R_t - \mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 f_R(x)\, dx.$$

Note: $m_1 = \mathbb{E}(R_t)$: the mean of the return distribution corresponds to the first non-central moment.
Note: $\sigma^2 = \mathbb{E}\left[(R_t - m_1)^2\right]$: the variance the return distribution corresponds to the second non-central moment.

Volatility

Definition (Volatility)

Volatility is defined as the standard deviation of the return distribution

$$\text{Volatility} = \sigma = \sqrt{\mathbb{V}\left(R_t\right)}.$$

- Volatility is generally expressed on an annual basis.
- Daily volatility computed from daily (or intraday returns) thus needs to be **annualized**.

Example (Annualized volatility)

Knowing that the standard deviation of daily returns in the period Jan. 4$^{\text{st}}$, 2010 to Dec. 31$^{\text{st}}$, 2018 for AAPL is 0.0130, determine the corresponding annualized volatility. Assume there are 252 business days in a year.

$$\text{Annualized Volatility} = \sqrt{252} \cdot 0.0130 = 0.2059 = 20.6\%$$

Skewness

Definition (Skewness)

The third central moment measures the **skewness** of the distribution

$$\mu_3 = \mathbb{E}\left[(R_t - \mu)^3\right] = \int_{-\infty}^{+\infty} (x - \mu)^3 f_R(x)\, dx.$$

The (standardized) skewness coefficient is defined as:

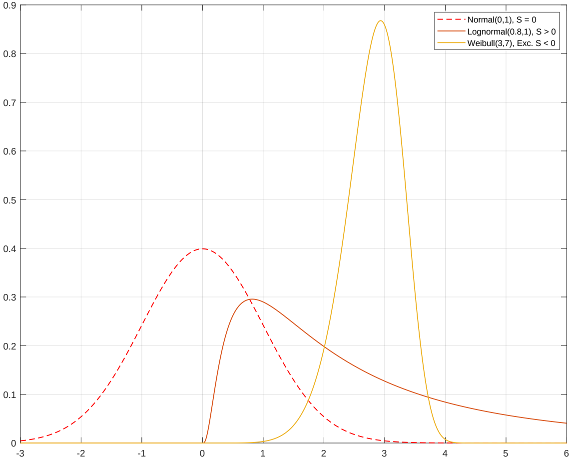
$$\mathbb{S}(R_t) = \mathbb{E}\left[\left(\frac{R_t - \mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}.$$

- Skewness captures symmetry of the distribution.

Skwness (cont'd)

- When $\mathbb{S}(R_t) < 0$, large realizations of R_t are more often negative than positive (e.g. market crashes are more likely than booms).
- With $\mathbb{S}(R_t) < 0$, the distribution is said to be "left-skewed", "left-tailed", or "skewed to the left".
- For symmetric distributions (e.g., Normal, Student, etc.), the skewness is null: $\mathbb{S}(R_t) = 0$.

Skwness (cont'd)



Kurtosis

Definition (Kurtosis)

The fourth central moment measures the **kurtosis** of the distribution

$$\mu_4 = \mathbb{E}\left[(R_t - \mu)^4\right] = \int_{-\infty}^{+\infty} (x - \mu)^4 f_R(x)\, dx.$$

The (standardized) kurtosis coefficient is defined as:

$$\mathbb{K}(R_t) = \mathbb{E}\left[\left(\frac{R_t - \mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4}.$$

- Kurtosis captures tail heaviness of the distribution.

Kurtosis (cont'd)

Kurtosis of the normal distribution

For a normal distribution $R_t \sim \mathcal{N}\left(\mu, \sigma^2\right)$,

$$\mathbb{K}\left(R_t\right) = 3.$$

Note: No matter what are μ and σ^2, when you compute

$$\int_{-\infty}^{+\infty} (x-\mu)^4 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

you always get 3!

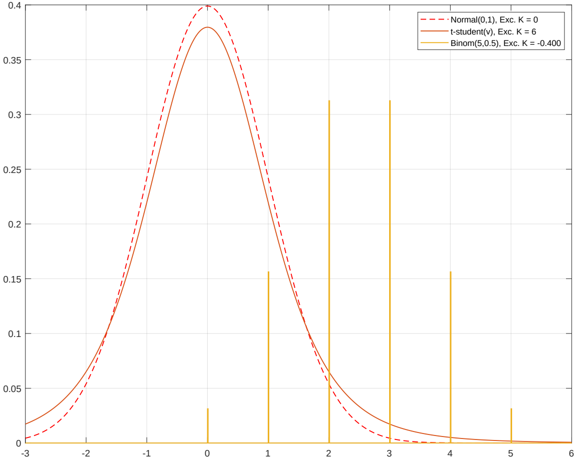
Note: The normal distribution is not the only distribution having kurtosis equal to 3.

Kurtosis properties

- Large $\mathbb{K}(R_t)$ implies that large realizations (positive or negative) are more likely to occur (compared to a normal distribution)
- $\mathbb{K}(R_t) > 3$, the distribution of R_t, is said to be **leptokurtic**.
- $\mathbb{K}(R_t) = 3$, the distribution of R_t, is said to be **mesokurtic**.
- $\mathbb{K}(R_t) < 3$, the distribution of R_t, is said to be **platykurtic**.

- The **excess kurtosis** is equal to $\mathbb{K}(R_t) - 3$.
  - Exc. Kurt > 0, stands for tails heavier than that of the normal distribution.
  - In general, mind that a visual inspection of the pdfs can be pointless in this regard.

Kurtosis (cont'd)



Recap

Each (central or non-central) moment provides information on the distribution of returns

	Formula	Interpretation
Mean	$\mathbb{E}\left(R_t\right)=\mu$	Indicator of central tendency
Variance	$\mathbb{V}\left(R_t\right)=\mathbb{E}\left[\left(R_t-\mu\right)^2\right]=\sigma^2$	Indicator of dispersion around μ
Skewness	$\mathbb{S}\left(R_t\right)=\mathbb{E}\left[\left(R_t-\mu\right)^3\right]/\sigma^3$	Indicator of symmetry
Kurtosis	$\mathbb{K}\left(R_t\right)=\mathbb{E}\left[\left(R_t-\mu\right)^4\right]/\sigma^4$	Indicator of tail heaviness

Moments of the normal distribution

Example

If we assume that returns are normally distributed with $R_t \sim \mathbb{N}\left(\mu, \sigma^2\right)$, then

$$
\begin{aligned}
\mathbb{E}\left(R_t\right) &= \mu \\
\mathbb{V}\left(R_t\right) &= \sigma^2 \\
\mathbb{S}\left(R_t\right) &= 0 \\
\mathbb{K}\left(R_t\right) &= 3
\end{aligned}
$$

The normal distribution is symmetric and mesokurtic.

Moments of the Student distribution

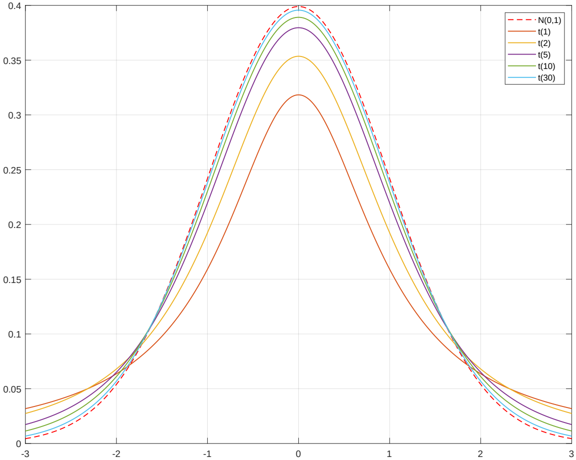
Example

If we assume that returns follow a Student-t distribution with ν degrees of freedom, $R_t \sim t(\nu)$, then

$$\mathbb{E}(R_t) = 0$$
$$\mathbb{V}(R_t) = \frac{\nu}{\nu - 2} \quad \text{if} \quad \nu > 2$$
$$\mathbb{S}(R_t) = 0 \quad \text{if} \quad \nu > 3$$
$$\text{Ex. } \mathbb{K}(R_t) = \frac{6}{\nu - 4} \quad \text{if} \quad \nu > 4$$

The Student-t distribution is generally **symmetric** and **leptokurtic**. Its excess kurtosis decreases as the number of degrees of freedom grows. When ν tends to ∞, $t(\nu) \approx \mathbb{N}(0, 1)$, and the distribution is mesokurtic.

Student to Normal convergence



Note: Convergence in distribution means $\lim_{n \to +\infty} F_n(x) = F(x)$

Remark

- Moments allow to characterize some aspects of the shape of the returns' distribution.
- However, the (theoretical) moments are unobservable: we need to **estimate** them.
- Denote by $\{R_1, \ldots, R_t\}$ a **random sample** of i.i.d. random variables that have the same distribution (R_t).
- Denote by $\{r_1, \ldots, r_t\}$ the realization of this sequence. This is your **dataset**.

Statistics

Definition

Let $X_1, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, \ldots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $(X_1, \ldots, X_n)$. Then the random variable or random vector $Y = T(X_1, \ldots, X_n)$ is called **statistic**. The probability distribution of statistic Y is called **sampling distribution** of Y.

Definition

A **sample statistic** is defined as any number computed from your sample data, i.e. $y = T(x_1, \ldots, x_n)$.

Population parameter

Definition

A **population parameter** (or just parameter) is defined as any number computed from the **entire** population (Siegel, 2016).

Remarks:

- A parameter is a fixed number because no randomness is involved. However, you will not usually have data available for the entire population. Therefore, a parameter is unknown and fixed (Siegel, 2016).
- There is often a natural correspondence between statistics and parameters. For each population parameter (a number you would like to know but cannot know exactly), there is a sample statistic computed from data that represents your best information about the unknown parameter.

Estimator

Definition

An estimator is a function of a sample related to some population parameter.

Remark:

- Any estimator is a statistics.
- However,
  - a statistic is a function of a sample.
  - an estimator is a function of a sample related to some parameter of relevance.
- The actual number computed from the data is called an **estimate** of the population parameter.

Example (Sample mean)

Assume $R_1, \ldots, R_T$ are i.i.d. rv. The sample mean

$$\hat{\mu}_T = \bar{R}_T = \frac{1}{T} \sum_{t=1}^{T} R_t,$$

is an estimator of the population mean $\mu = \mathbb{E}\left(R_t\right)$.

Examples (cont'd)

Example (Sample variance)

Assume $R_1, \ldots, R_T$ are i.i.d. rv. The sample variance

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

is an estimator of the population variance $\sigma^2 = \mathbb{V}(R_t)$.

Examples (cont'd)

Example (Sample skewness)

Assume $R_1, \ldots, R_T$ are i.i.d. rv. The sample skewness

$$\hat{S}_T = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{R_t - \bar{R}_T}{\hat{\sigma}}\right)^3 = \frac{\frac{1}{T}\sum_{t=1}^{T}\left(R_t - \bar{R}_T\right)^3}{\left(\frac{1}{T-1}\sum_{t=1}^{T}\left(R_t - \bar{R}_T\right)^2\right)^{\frac{3}{2}}}$$

is an estimator of the population skewness $\mathbb{S}\left(R_t\right)$.

Examples (cont'd)

Example (Sample kurtosis)

Assume $R_1, \ldots, R_T$ are i.i.d. rv. The sample kurtosis

$$\hat{K}_T = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{R_t - \bar{R}_T}{\hat{\sigma}}\right)^4 = \frac{\frac{1}{T}\sum_{t=1}^{T}\left(R_t - \bar{R}_T\right)^4}{\left(\frac{1}{T-1}\sum_{t=1}^{T}\left(R_t - \bar{R}_T\right)^2\right)^2}$$

is an estimator of the population kurtosis $\mathbb{K}\left(R_t\right)$.

Question: why do we have in the denominator $T-1$ (both for skewness and kurtosis)? What if we divide by T? Can we do this?

Sampling distribution

Fact

An estimator $\hat{\theta}$ for a parameter θ is a **random variable**.

Consequence: $\hat{\theta}$ has a (marginal or conditional) **probability distribution**. This **sampling distribution** is characterized by a pdf $f_{\hat{\theta}}$.

Definition (Sampling distribution)

The probability distribution of an estimator (or a statistic) is called the sampling distribution.

Consequence: The sampling distribution of $\hat{\theta}$ is characterized by moments such as mean $\mathbb{E}(\hat{\theta})$, variance $\mathbb{V}(\hat{\theta})$, etc.

Estimate

Definition (Point estimate)

A (point) estimate is the realized value of an estimator (i.e. a number) that is obtained when a sample is actually taken. For an estimator $\hat{\theta}$ it can be denoted by $\hat{\theta}\,(data)$.

Example (Point estimate)

For instance a realization of he estimator $\bar{R}_T$, denoted by $\bar{r}_t$, is defined as

$$\bar{r}_T = \frac{1}{T} \sum_{t=1}^{T} r_t$$

- if $T = 3$ and $\{r_1, r_2, r_3\} = \{1, 2, 3\}$ then $\bar{r}_T = 2$.
- if $T = 3$ and $\{r_1, r_2, r_3\} = \{-4, 0, 8\}$ then $\bar{r}_T = 4/3$.

Properties of estimators

Question: what constitutes a good estimator?

Properties of estimators (cont'd)

Estimators are compared on the basis of a variety of attributes.

- **Finite sample** properties (or finite sample distribution) of estimators are those attributes that can be compared regardless of the sample size.
- However, the finite sample distribution is known only in the case of specific distributional assumption on R_t (typically normality).
- When the normality assumption is not applicable (and likely the finite sample distribution is unknown), estimators are evaluated on the basis on their large sample, or **asymptotic properties**.

Finite sample distribution

Theorem

If we assume that the returns $R_1, \ldots, R_T$ are i.i.d. with $R_t \sim \mathcal{N}\left(\mu, \sigma^2\right)$, then $\hat{\mu}_T$ and $(T-1)\,\hat{\sigma}^2_T/\sigma^2$ have finite sampling distribution

$$\hat{\mu}_T \sim \mathcal{N}\left(\mu, \sigma^2/T\right) \quad \forall T \in \mathbb{N}$$
$$(T-1)\,\hat{\sigma}^2_T/\sigma^2 \sim \chi\,(T-1) \quad \forall T \geq 2$$

Example

Consider $R_t \sim \mathcal{N}\left(\mu, \sigma^2\right), t = 1, \ldots, 10$, then

$$\hat{\mu}_{10} = \mathcal{N}\left(\mu, \frac{\sigma^2}{10}\right) \qquad 9\frac{\hat{\sigma}^2_T}{\sigma^2} \sim \chi\,(9)$$

Sampling distribution of the sample mean, proof

- **Sampling distribution** of the sample mean

  Let $X_1, \ldots, X_T$ be i.i.d. normal r.v. with mean μ and variance σ^2, take the sample mean estimator:

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

  Fact: a linear combination of normal r.v. is normal, thus $\hat{\mu} \sim \mathcal{N}$.

Question: How that the mean and variance of $\hat{\mu}$ are respectively μ and σ^2/T?
Hint: use the properties of the expectation and variance to compute $\mathbb{E}(\hat{\mu})$ and $\mathbb{V}(\hat{\mu})$, and keep in mind that the rv are i.i.d.

Finite sample distribution (cont'd)

In most of cases, it is impossible to derive the exact/finite sample distribution for the estimator (or a transformed variable).
Two reasons:

1. the exact distribution of $R_1, \dots, R_T$ is known, but the function $T(\cdot)$ is too complicated to derive the distribution of $\hat{\theta}$.

$$T(R_1, \dots, R_T) \sim ???$$

2. the distribution of $R_1, \dots, R_T$ is unknown.

Asymptotic theory

Question: what is the behavior of the estimator $\hat{\theta}_T$ when the sample size T tends to infinity?

Definition (Asymptotic theory)

Asymptotic or large sample theory consists in the study of the distribution of the estimator when the sample size is large.

The asymptotic theory is fundamentally based on the notion of **convergence**.

Convergence

For our purposes, we focus on two types of convergence

1. **Convergence in probability** (or almost sure):

$$\hat{\theta}_T \text{ converges to a real number}$$

   This convergence types are used to derive the **consistency** property of the estimators.

2. **Convergence in distribution**:

$$\sqrt{T}\left(\hat{\theta}_T - \theta\right) \text{ converges to a given distribution}$$

   This convergence is used to derive the asymptotic distribution of the estimators and to make inference (tests) about the true value of the parameters.

Convergence in probability

Definition (Convergence in probability)

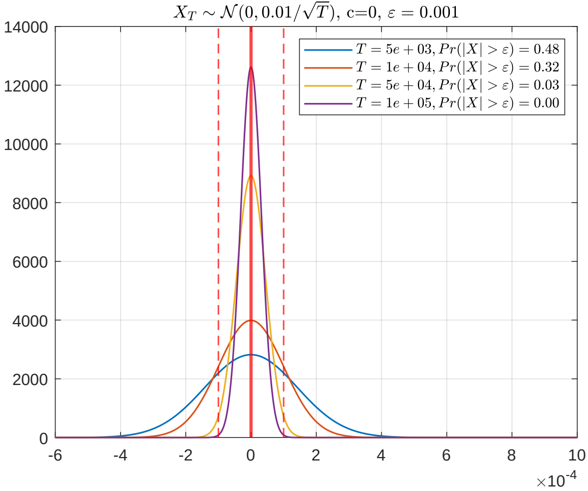
Let X_T be a sequence of random variables indexed by the sample size. X_T converges in probability to a constant c, if, for any $\varepsilon > 0$

$$\lim_{T \to \infty} \Pr\left(|X_T - c| > \varepsilon\right) = 0.$$

It is written

$$X_T \xrightarrow{p} c \quad \text{or} \quad \operatorname{plim} X_T = c.$$

Consistency (cont'd)



Consistency

Theorem (Consistency)

If returns are i.i.d. with finite moments $\mathbb{E}(R_t)$, $\mathbb{V}(R_t)$, $\mathbb{S}(R_t)$ *and* $\mathbb{K}(R_t)$*, then the sample first four moments are* **weakly consistent***:*

$$\hat{\mu}_T \xrightarrow{p} \mathbb{E}(R_t) = \mu$$

$$\hat{\sigma}_T^2 \xrightarrow{p} \mathbb{V}(R_t) = \sigma^2$$

$$\hat{S}_T \xrightarrow{p} \mathbb{S}(R_t)$$

$$\hat{K}_T \xrightarrow{p} \mathbb{K}(R_t)$$

Consistency (cont'd)

Interpretation:

1. The distribution of the sample moments is highly concentrated around the true value (unknown) of the population moments of the returns when the sample size T tends to infinity.
2. The realization of the sample moments are then 'close' to the value of the population moment when the sample size T tends to infinity.

Consistency of the sample mean, proof

- **Consistency** of the sample mean.

  Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2, take the sample mean estimator:

  $$\mathbb{E}(\hat{\mu}) = \mu, \qquad \mathbb{V}(\hat{\mu}) = \frac{\sigma^2}{T}$$

  Since we do not know the distribution of X_t we do not know the pdf of $\hat{\mu}$. All we know is that $\mathbb{E}(\hat{\mu}) = \mu$ and $\mathbb{V}(\hat{\mu}) = \frac{\sigma^2}{T}$. However, as $n \to \infty$, $\mathbb{V}(\hat{\mu}) \to 0$ and the pdf of $\hat{\mu}$ collapses to μ. Intuitively, as $n \to \infty$, $\mathbb{E}(\hat{\mu})$ converges in some sense to μ. In other words, the estimator $\hat{\mu}$ is consistent for μ.

Note: Consistency for the sample variance is simple to proof but requires tedious and long algebra.

Convergence in distribution

Definition (Convergence in distribution)

Let X_T be a sequence random variable indexed by the sample size with a cdf $F_T(\cdot)$. X_T converges in distribution to a random variable X with cdf $F(\cdot)$ if

$$\lim_{T\to\infty} F_T(x) = F(x) \quad \forall x.$$

It is written:

$$X_T \xrightarrow{d} X$$

Asymptotic distributions for sample mean and variance

Theorem (Asymptotic distributions for sample mean and variance)

If returns $R_1, \ldots, R_T$ are i.i.d. with finite mean $\mathbb{E}(R_t) = \mu$ and finite variance $\mathbb{V}(R_t) = \sigma^2$, we can derive the following **asymptotic distributions** *for T that grows to infinity:*

$$\sqrt{T}\left(\hat{\mu}_T - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$$
$$\sqrt{T}\left(\hat{\sigma}^2_T - \sigma^2\right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^4\right)$$

Question: does the specific distribution for R_t matter?

Convergence in distribution (cont'd)

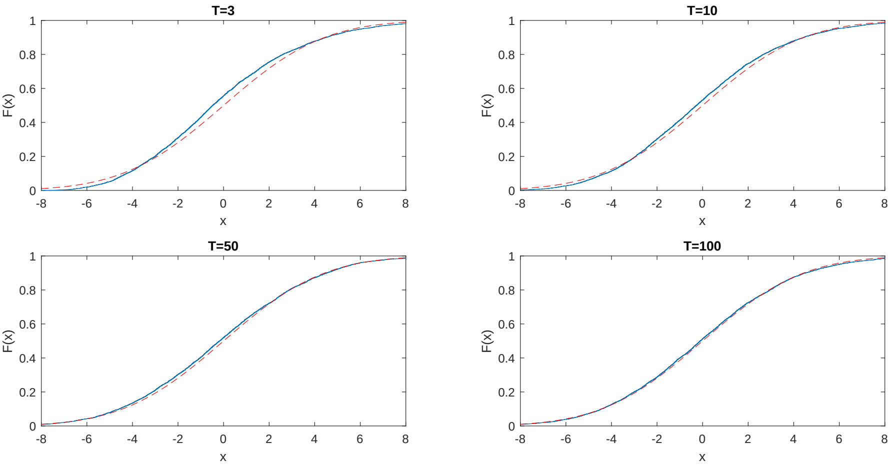


Figure: ECDF of 10^5 realizations of $\sqrt{T}\,(\hat{\mu}_T - 6)$ from a sample of size T drawn from a Gamma(3,2) against a $\mathcal{N}(0, 12)$.

Asymptotic dist. for sample mean and variance (cont'd)

Remark: the previous result can be interpreted as

$$\hat{\mu}_T \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{T}\right) \qquad \hat{\sigma}_T^2 \approx \mathcal{N}\left(\sigma^2, \frac{2\sigma^4}{T}\right)$$

with $\approx$ meaning 'asymptotically distributed as'.

Note: These asymptotic results for the sample moments can be used to perform **statistical tests** about the (population) moments.

Example

Example

Consider a sample of $T = 500$ i.i.d. returns $R_1, \ldots, R_T$, and a realization (dataset) $r_1, \cdots, r_T$ with

$$\hat{\mu}_T = -0.05 \qquad \hat{\sigma}_T^2 = 0.0001.$$

Question: compute a 95% confidence interval on the true vale of the expected return $\mathbb{E}(R_t)$.

Example (cont'd)

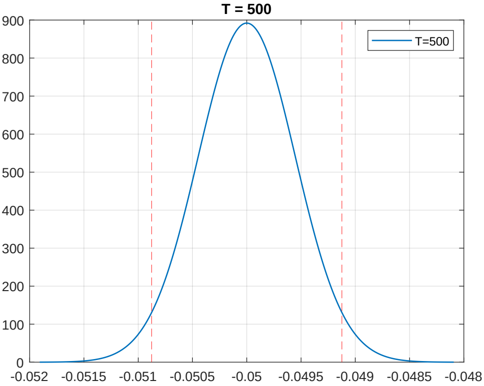
Answer: We know that $\hat{\mu}_T \approx \mathcal{N}\left(\mu, \sigma^2/T\right)$, so we have:

$$IC_{95\%} = \left[\mu \pm \frac{\sigma}{\sqrt{T}}\Phi^{-1}(0.975)\right].$$

Since μ and σ are unobservable, we use a consistent estimator (plug-in approach) and compute:

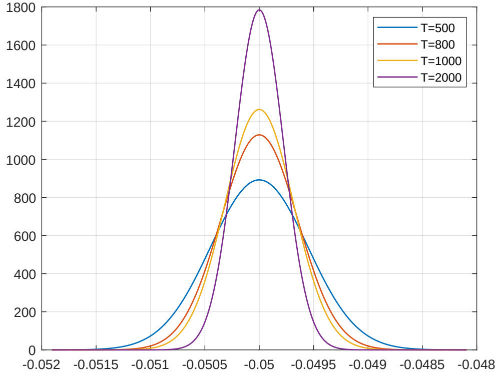
$$\begin{aligned}
IC_{95\%} = & \left[\mu \pm \frac{\sigma}{\sqrt{T}}\Phi^{-1}(0.975)\right] \\
& \left[\hat{\mu} \pm \frac{\hat{\sigma}}{\sqrt{T}}\Phi^{-1}(0.975)\right] \\
& \left[-0.05 \pm \sqrt{\frac{0.0001}{500}}\Phi^{-1}(0.975)\right] \\
& [-0.0509 \; - 0.0491].
\end{aligned}$$

Example (cont'd)



Example (cont'd)

Question: what happens if the same values of the sample mean and variance are observed over increasing sample sizes?



Asymptotic dist. for sample mean, proof

Theorem (Lindberg-Levy Central limit theorem)

Let $X_1, \ldots, X_T$ be an i.i.d. sample with $\mathbb{E}(X_t) = \mu$ and $\mathbb{V}(X_i) = \sigma^2 < \infty$. Then (Greene, 2003, p. 909),

$$Y_T = \sqrt{T}\left(\frac{\bar{X} - \mu}{\sigma}\right) \xrightarrow{d} \sim \mathcal{N}(0, 1) \quad \textit{as } n \to \infty.$$

Remark: the CLT suggests that we may approximate the distribution of $Y_T = \sqrt{T}\left(\frac{\bar{X}-\mu}{\sigma}\right)$ by a standard normal distribution. This, in turn, suggests approximating the distribution of (any) sample average $\bar{X}$ by a normal distribution.

Asymptotic dist. for sample skewness and kurtosis

Theorem

If we assume that the returns $R_1, \ldots, R_T$ are $\mathcal{N}\left(\mu, \sigma^2\right)$, then we have (Snedecor and Cochran, 1980, p. 78):

$$\sqrt{T}\left(\hat{S}_T - 0\right) \xrightarrow{d} \mathcal{N}\left(0, 6\right),$$

$$\sqrt{T}\left(\hat{K}_T - 3\right) \xrightarrow{d} \mathcal{N}\left(0, 24\right).$$

Note: These asymptotic results for the sample moments can be used to perform statistical tests about the returns distribution (Jarque-Bera test for instance).

Jarque-Bera test

- Given the asset return series $r = \{r_1 \dots r_T\}$ we consider the null hypothesis $H_0 : \mathbb{S}(r) = 0$, versus the alternative $H_1 : \mathbb{S}(r) \neq 0$. The t-ratio statistics of the sample skewness is:

$$t = \frac{\hat{\mathbb{S}}(r)}{\sqrt{6/T}}$$

- Similarly one can test the excess kurtosis using the hypotheses $H_0 : \mathbb{K}(r) - 3 = 0$, versus $H_1 : \mathbb{K}(r) - 3 \neq 0$. The test statistics is:

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

The decision rules go as follows. Reject the null at as significance level α if $|t| > Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is the upper $100 \times (1 - \alpha/2)$ quantile of the standard normal distribution.

Jarque-Bera test (cont'd)

Jarque and Bera, 1987 (JB) combine the two prior tests and test the normality of the return distribution by using the test statistics

$$\mathrm{JB} = \frac{\hat{\mathbb{S}}^2(r)}{6/T} + \frac{\left[\hat{\mathbb{K}}(r) - 3\right]^2}{24/T},$$

for which

$$\mathrm{JB} \xrightarrow{d} \chi^2(2).$$

One rejects H_0 of normality if the p-value of the JB statistics is less than a set significance level.

Question: why $\chi^2(2)$?

[1] Carlos M Jarque and Anil K Bera. 'A test for normality of observations and regression residuals'. In: *International Statistical Review/Revue Internationale de Statistique* (1987), pp. 163–172

Distribution of returns

A **traditional assumption** is that simple returns $\{R_{t,\, t=1,\dots,T}\}$ are i.i.d. **normal** with fixed mean and variance.

+ Statistical properties of asset returns are tractable.
- The lower bound for a simple return is -1, but the normal distribution has no lower bound.
- Multiperiod returns are not normally distributes as the product of one-period returns is not normal.

Distribution of returns (cont'd)

A **common assumption** is that logreturns $\{r_t \; t = 1, \ldots, T\}$ are i.i.d. normal with fixed mean μ and variance σ^2. The simple returns are then i.i.d. lognormal rv with mean m_1 and variance m_2

$$m_1 = \exp\left(\mu + \frac{\sigma^2}{2}\right) - 1, \quad m_2 = \exp(2\mu + \sigma^2)\left[\exp\left(\sigma^2\right) - 1\right]$$

+ The two equations are useful for forecasting models built for log-returns.
+ The multiperiod logreturn is normally distributed.
+ There is no lower bound for r_t.
- The lognormal assumption is not consistent with empirical properties of historical stock returns (e.g. excess kurtosis).

Distribution of returns (cont'd)

Other options for logreturns include:

- **Stable distributions**: (+) linear combinations are still stable distributions, (+) reproduce excess kurtosis, (-) difficult to use, (-) non-normal stable distributions have infinite variance (e.g. Cauchy distribution).
- **Scale mixture of normal distribution**: $r_t \sim \mathcal{N}\left(\mu, \sigma^2\right)$ but σ^2 is a r.v. that follows a positive distribution (e.g. σ^{-2} follows a gamma distribution).
- **Finite mixture of normal distributions**:

$$r_t \sim (1 - X)\,\mathcal{N}\left(\mu, \sigma_1^2\right) + X\,\mathcal{N}\left(\mu, \sigma_2^2\right)$$

  with X is a Bernoulli r.v. such that $\Pr(X = 1) = \alpha$, σ_1^2 is small, and σ_2^2 is relatively large.

Volatility process

Note:

- So far we focused on the return process.
- As well, one might study other processes such as durations, volumes, bid-ask spread etc.
- Volatility plays a central role in risk management and derivative pricing. The study of the **volatility process** - concerned with the evolution of the conditional variance of return over time - is one of the major topics in financial econometric

Section 5, Stylized facts

Stylized facts

- The statistical properties of financial data revealed a wealth of stylized facts.
- These stylized facts are common in a wide range of financial time series.

[1] Rama Cont. 'Empirical properties of asset returns: stylized facts and statistical issues'. In: *Quantitative finance* 1.2 (2001), p. 223

Stylized facts (cont'd)

(Fan and Yao, 2017) identify eight main stylized facts:

1. Stationarity
2. Absence of autocorrelations
3. Asymmetry
4. Volatility clustering
5. Aggregational gaussianity
6. Long-range dependence
7. Leverage effects

[1] Jianqing Fan and Qiwei Yao. *The elements of financial econometrics*. Cambridge University Press, 2017

Stylized fact 1, stationarity

Fact 1, stationarity

In general, prices are non-stationary whereas returns are **stationary**.

- The prices of an asset recorded over times are often not stationary due to, for example, the steady expansion of economy, the increase of productivity, and financial crisis.
- However their returns, typically fluctuates around a constant level, suggesting a constant mean over time.
- Most return sequences can be modeled as a stochastic processes with at least time-invariant two fist moments: **weak stationarity**.

Weak stationarity

Definition

Weak or second-order stationarity A time-series process $(X_t, t \in \mathbb{Z})$ is **weakly stationary** (second-order stationary) if and only if:

- $\forall t \in \mathbb{Z}, \mathbb{E}\left[X_t^2\right] < \infty$
- $\forall t \in \mathbb{Z}, \mathbb{E}\left[X_t\right] = m$ is not dependent on t
- $\forall (t, h) \in \mathbb{Z}^2, \mathbb{C}ov\left(X_t, X_{t+h}\right) = \mathbb{E}\left[\left(X_{t+h} - m\right)\left(X_t - m\right)\right] = \gamma_h$ does not depend on t

Stylized fact 1, stationarity (cont'd)

Stylized fact 2, absence of autocorrelations

Fact 2, absence of autocorrelations

The **autocorrelations** of asset returns are often insignificant, except for small intraday time scales (approx. less than 5 minutes) for which microstructure effects come into play.

Note: The fact that returns hardly show any serial correlation, does not mean that they are independent.

Autocorrelation

Definition (Autocorrelation)

The **autocorrelation**, denoted ρ_k , of a weak stationary process R_t is the correlation between values of the process at different times, defined as:

$$\rho_k = \mathbb{C}orr\left(R_t, R_{t-k}\right) = \frac{\mathbb{E}\left[\left(R_t - \mu\right)\left(R_{t-k} - \mu\right)\right]}{\mathbb{V}\left(R_t\right)} = \frac{\gamma_k}{\sigma^2}$$

with $\mu = \mathbb{E}\left(R_t\right)$, $\sigma^2 = \mathbb{V}\left(R_t\right)$, $\forall t$. γ_k is the autocovariance of order k.

Sample autocorrelation

Definition (Sample autocorrelation)

The **autocorrelation**, denoted $\hat{\rho}_k$, of a weak stationary process R_t is the estimator or ρ_k defined as

$$\hat{\rho}_k = \frac{1}{(T-k)\,\hat{\sigma}^2} \sum_{t=k+1}^{T} (r_t - \hat{\mu})(r_{t-k} - \hat{\mu})$$

with $\hat{\mu}$ and $\hat{\sigma}^2$ are consistent estimators of $\mu = \mathbb{E}(R_t)$ and $\sigma^2 = \mathbb{V}(R_t)$, $\forall t$.

White noise

Definition (White noise process)

A **white noise** process is a random process of random variables that are uncorrelated, have mean zero, and a finite variance. $X_t \sim WN\left(0, \sigma^2\right)$ process if

- $\mathbb{E}\left(X_t\right) = 0$
- $\mathbb{E}\left(X_t^2\right) = \sigma^2$
- $\mathbb{E}\left(X_t X_h\right) = 0, \forall t \neq h$

Note: by definition its autocorrelation is zero.

ACF

The Autocorrelation function (ACF) (or correlogram) represents the sample autocorrelation at different lags from $k = 1$ (or 0) to some maximum lag order, say $k = 15$.

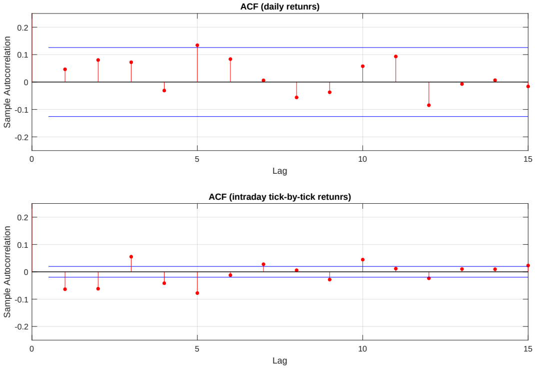


Figure: Sample ACF for INTC. Top: 2011, bottom: Dec. 27, 2018.

ACF (cont'd)

Testing for **individual** significance:

- Under the assumption that $R_t \sim WN(0, \sigma^2)$, the distribution of ρ_k in large samples is approximated by

$$\sqrt{T}\rho_j \sim \mathcal{N}(0, 1).$$

- **Decision rule**:
  for any $k = 1, 2, \ldots, T$ we reject the null $H_0 : \rho_k = 0$ in favor of the alternative hypothesis $H_1 : \rho_k \neq 0$ at at a significance level α if:

$$|\rho_j| > \Phi^{-1}(1 - \alpha/2)/\sqrt{T}$$

  where $\Phi^{-1}(\cdot)$ is the inverse of the normal cdf (i.e. quantile function).

ACF (cont'd)

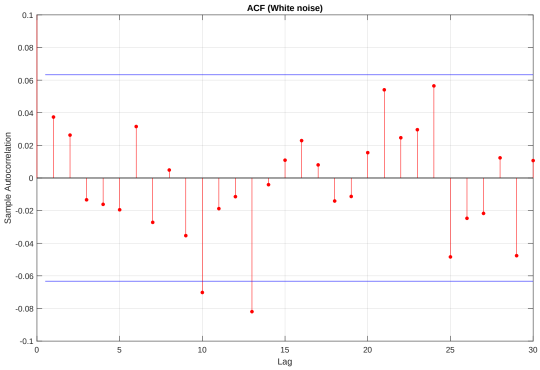


Figure: Sample ACF for a white noise sample

Question: how that for lags 10 and 13 sample autocorrelationa are beyond the 95% CI?

Q-test

Testing the nullity of autocorrelations

- It is possible to test the nullity of the autocorrelations through a **Q-test** (Box-Pierce test or Ljung-Box test).
- For a given lag-order K, the null hypothesis of the test is

$$H_0 \; : \; \rho_1 = \cdots = \rho_K = 0.$$

- The alternative hypothesis is

$$H_1 \; : \; \exists k \in \{1, \ldots, K\} \text{ s.t. } \rho_k \neq 0$$

Q-tests (cont'd)

Box-Pierce test

The Box-Pierce test statistic associated to the null H_0 is defined as

$$Q_{BP} = T \sum_{k=1}^{K} \hat{\rho}_k^2 \xrightarrow[H_0]{d} \chi_K^2.$$

Decision rule

- If the realization of the test statistic Q_{BP} is larger than the quantile of a chi-squared distribution with K degrees of freedom at the probability $1 - \alpha\%$ (say, 95%), for a risk level $\alpha\%$ (say, 5%) we reject the null hypothesis H_0.
- If the p-value associated to the test statistic Q_{BP} is smaller than the risk level $\alpha\%$, we reject the null hypothesis H_0.
- If the null hypothesis H_0 is rejected, there is at least one autocorrelation (between lag 1 and k) which is non-null, i.e. the time series R_t is autocorrelated.

Q-tests (cont'd)

	Daily (2011)			Intraday (Dec. 27, 2018)		
Lag	Reject	Statistics	P-val	Reject	Statistics	P-val
1	0	0.55	0.457	1	41.94	0.942E-11
2	0	2.21	0.331	1	81.60	0.000
3	0	3.55	0.314	1	113.01	0.000
4	0	3.80	0.433	1	130.97	0.000
5	0	8.49	0.131	1	193.44	0.000
6	0	10.32	0.112	1	194.97	0.000
7	0	10.33	0.171	1	202.87	0.000
8	0	11.16	0.193	1	203.21	0.000
9	0	11.53	0.241	1	211.55	0.000
10	0	12.41	0.258	1	232.04	0.000
11	0	14.73	0.195	1	233.42	0.000
12	0	16.64	0.164	1	239.28	0.000
13	0	16.66	0.216	1	240.35	0.000
14	0	16.67	0.274	1	241.27	0.000
15	0	16.74	0.335	1	246.71	0.000

Table: Box-Pierce test for INTC returns.

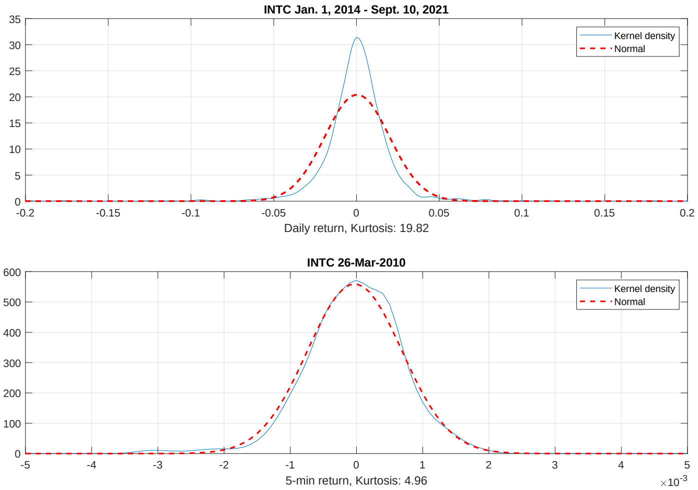
Question: How does this compare wrt. the previous ACF plot?

Stylized fact 3, heavy tails

Fact 3, heavy tails

The probability distribution of return often exhibits **heavier tails** than those of a normal distribution.

Stylized fact 3, heavy tails (cont'd)

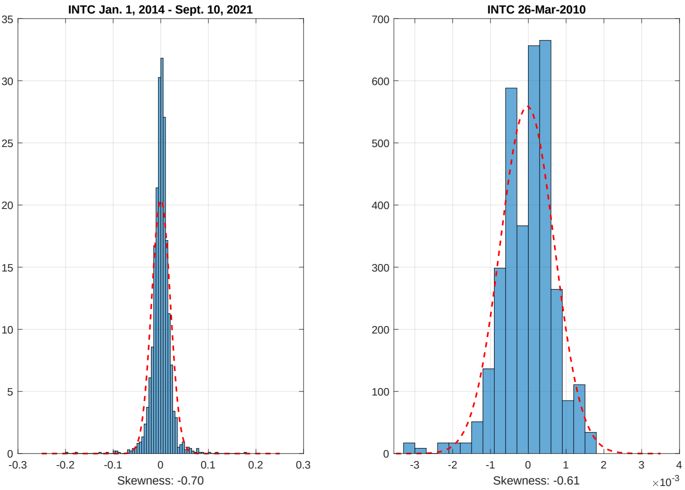


Stylized fact 4, asymmetry

Fact 4, asymmetry

The distribution of returns is **asymmetric** and often negatively skewed, reflecting the fact that the downturns of financial markets are often much steeper than the recoveries. Investors tend to react more strongly to negative news than to positive news.

Stylized fact 4, asymmetry (cont'd)



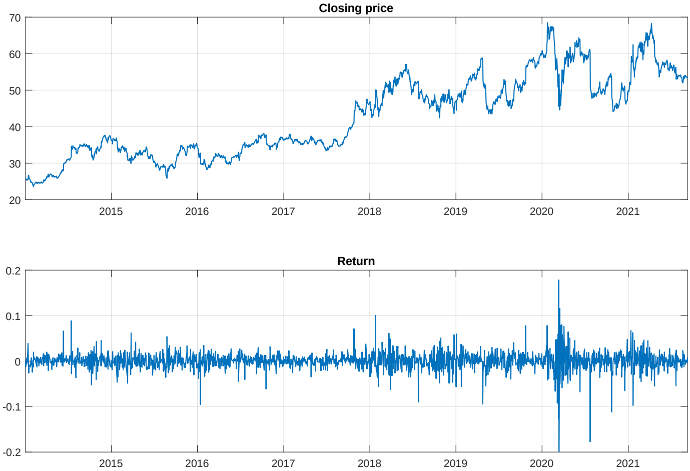
Stylized fact 5, volatility clustering

Fact 5, volatility clustering

The **volatility clustering** means that large price changes (i.e. returns with large absolute values or large squares) occur in clusters. Indeed, large price changes tend to be followed by large price changes, and periods of tranquility alternate with periods of high volatility.

Note: Note: the volatility clustering is the consequence of the autocorrelation of the squared returns (cf. stylized fact 7).

Stylized fact 5, volatility clustering (cont'd)



Stylized fact 5, volatility clustering (cont'd)

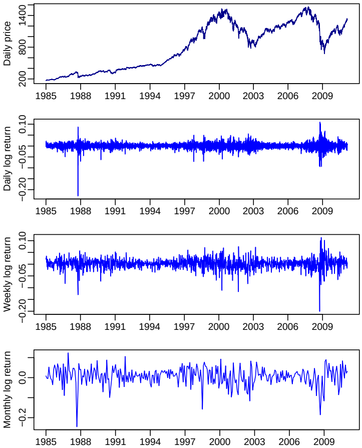


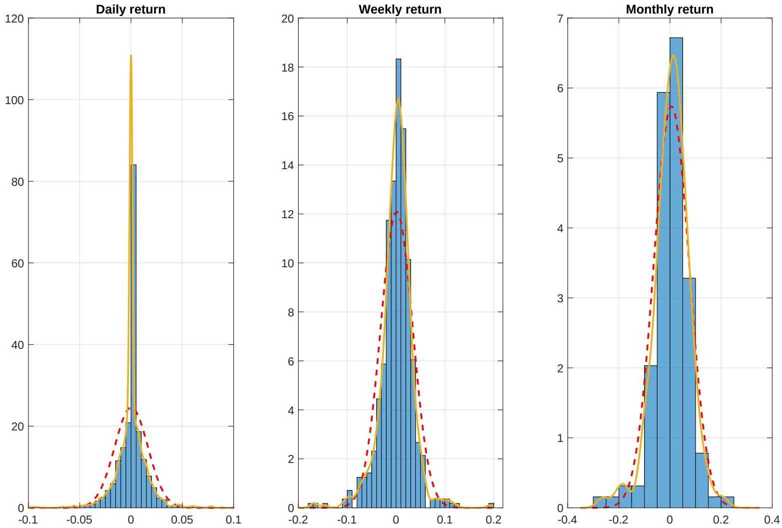
Figure: S&P 500 January 1985 - February 2011 (Fan and Yao, 2017).

Stylized fact 6, aggregational gaussianity

Fact 6, aggregational gaussianity

A return over k days is simply the aggregation of k daily returns. When the time horizon k increases, the central limit law sets in and the distribution of the returns over a long time-horizon (such as a month) tends toward a **normal distribution**.

Stylized fact 6, aggregational gaussianity (cont'd)



Stylized fact 6, aggregational gaussianity (cont'd)

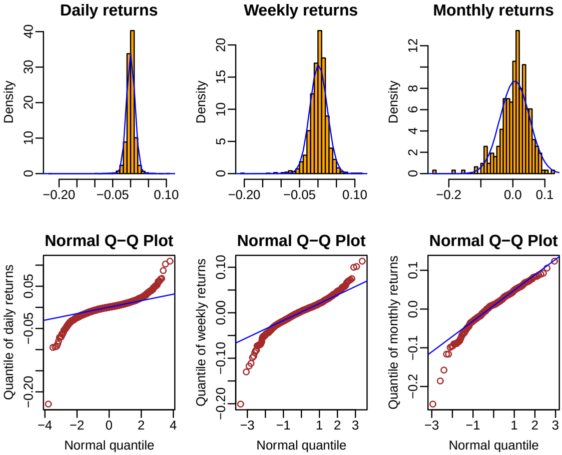


Figure: S&P 500 in January 1985 - February 2011 (Fan and Yao, 2017).

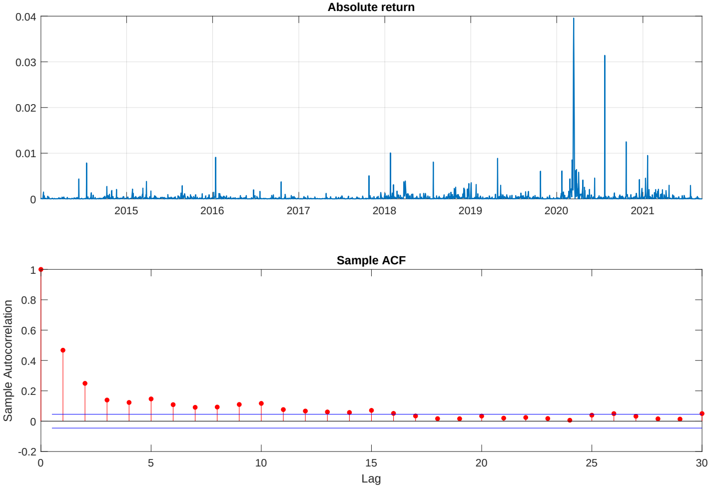
Stylized fact 7, long-range dependence

Fact 7, long-range dependence

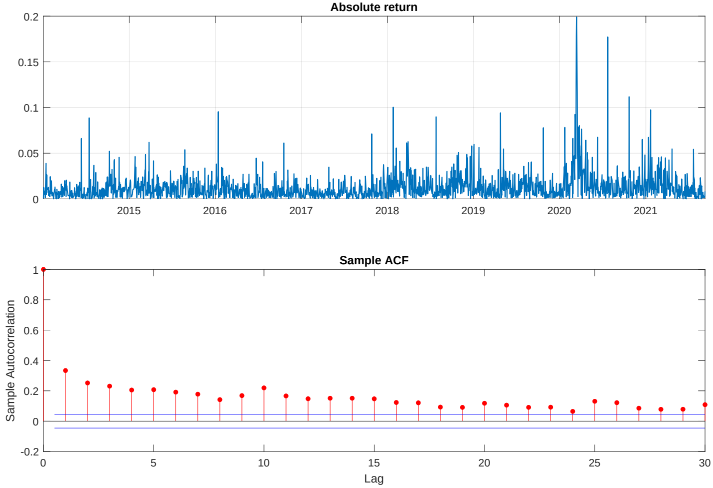
Both daily squared and absolute returns often exhibit significant autocorrelations. Those autocorrelations are persistent, indicating possible **long-memory** properties.

Note: Those autocorrelations become weaker and less persistent when the sampling interval is increased from a day, to a week to a month but are uqiquitios in high-frequency domains.

Stylized fact 7, long-range dependence (cont'd)



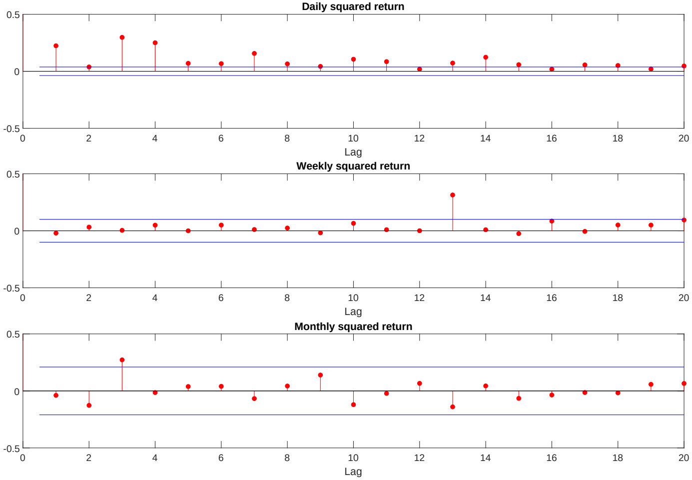
Stylized fact 7, long-range dependence (cont'd)



ARCH effects

- The autocorrelation of the squared returns is called **ARCH effect**.
- The ARCH effect depends on the sampling frequency.
- It is most important with daily returns, and less important with low frequency returns (monthly, quarterly, etc.).

ARCH effects (cont'd)



ARCH effects (cont'd)

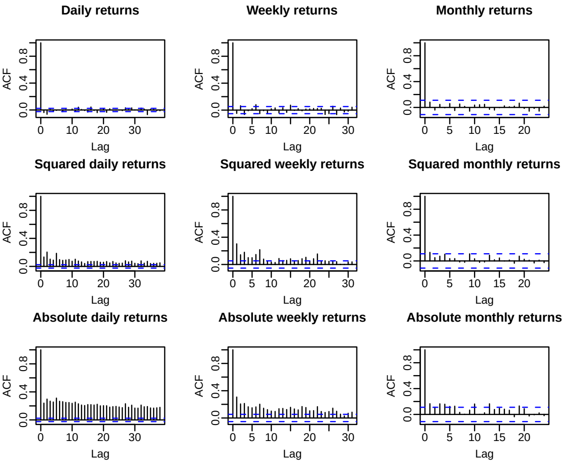


Figure: S&P 500 in January 1985 - February 2011 (Fan and Yao, 2017).

Stylized fact 8, leverage effect

Fact 8, leverage effect

Asset returns are negatively correlated with the changes of their volatilities: this negative correlation is called **leverage effect**.

- As asset prices decline, companies become more leveraged (debt to equity ratios increase) and riskier, and hence their stock prices become more volatile.
- On the other hand, when stock prices become more volatile, investors demand high returns and hence stock prices go down.
- Volatilities caused by price decline are typically larger than the appreciations due to declined volatilities.

Stylized fact 8, leverage effect (cont'd)

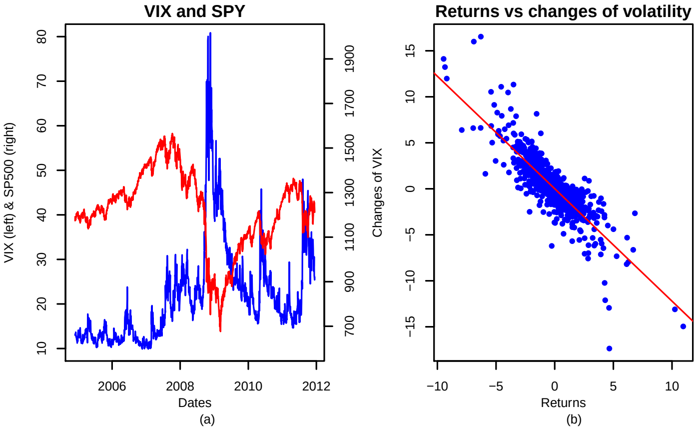


Figure: Time series plot of VIX (blue) and the S&P 500 index (red) in Nov. 29, 2004 - Dec. 14, 2011 (the left panel), and the plot of the daily S&P 500 returns (in percent) against the changes of VIX (the right panel) (Fan and Yao, 2017).

Section 6, References

Slides

Disclaimer:

- Some slides from Chirs Brooks' book (Brooks, 2014) (copyrighted)
- Some slides from Christophe Hurlin's (University of Orleans), financial econometrics course (2019), available online.
- Some slides original (made ad-hoc for this course)

Bibliography

[Fri33] Ragnar Frisch. 'Editor's Note'. In: *Econometrica* 1.1 (1933), pp. 1–4.

[SKS54] Paul A. Samuelson, Tjalling C. Koopmans and J. Richard N. Stone. 'Report of the evaluative committee for Econometrica'. In: *Econometrica* 22.2 (1954), pp. 141–146.

[SC80] George W Snedecor and William G Cochran. 'Statistical Methods Iowa State University Press, Ames'. In: *Statistical methods, 7th ed.. The Iowa State University Press, Ames* (1980).

[JB87] Carlos M Jarque and Anil K Bera. 'A test for normality of observations and regression residuals'. In: *International Statistical Review/Revue Internationale de Statistique* (1987), pp. 163–172.

Bibliography (cont'd)

[CLM96] John Y. Campbell, Andrew W. Lo and A. Craig MacKinlay. *The econometrics of financial markets*. Princeton University press, 1996.

[Con01] Rama Cont. 'Empirical properties of asset returns: stylized facts and statistical issues'. In: *Quantitative finance* 1.2 (2001), p. 223.

[Gre03] William H Greene. *Econometric analysis*. Pearson Education India, 2003.

[Fan04] Jianqing Fan. 'An introduction to financial econometrics'. In: *Department of Operation Research and Financial Engineering, Princeton University, Princeton* (2004).

[AY08] Yacine Aït-Sahalia and Jialin Yu. *High frequency market microstructure noise estimates and liquidity measures*. Tech. rep. National Bureau of Economic Research, 2008.

Bibliography (cont'd)

[CST10] Rama Cont, Sasha Stoikov and Rishi Talreja. 'A stochastic model for order book dynamics'. In: *Operations research* 58.3 (2010), pp. 549–563.

[Bro14] Chris Brooks. *Introductory Econometrics for Finance*. 3rd ed. Cambridge University Press, 2014. DOI: 10.1017/CBO9781139540872.

[Sie16] Andrew F Siegel. *Practical business statistics*. Academic Press, 2016.

[FY17] Jianqing Fan and Qiwei Yao. *The elements of financial econometrics*. Cambridge University Press, 2017.