Core Concepts in Financial Econometrics

Romain Lafarguette, Ph.D.

ADIA Quant & IMF External Consultant

Singapore Training Institute, 08 November 2022

Outline

This short crash-course is divided in three parts covering:

1. **Data concepts**: population, sample, data types, data generating process, etc.
2. **Statistical Inference**: stochastic process, estimator, distributions, convergence, etc.
3. **Statistical Properties of Financial Time Series**: stationarity, autocorrelation, heavy tails, etc.

NB: this slide-deck follows the excellent course of Christophe Hurlin
https://sites.google.com/view/christophe-hurlin/teaching-resources

Overview

Financial econometrics (including time-series econometrics) are based on four main elements:

1. A sample of data
2. An econometric model, based on a theory or not
3. An estimation method to estimate the coefficients of the model
4. Inference/testing approach to validate the estimation

Population vs. Sample

Definition: Population

A **population** is defined as including all entities (e.g. banks or firms) or all the time periods of the processus that has to be explained

- In most cases, it is impossible to observe the entire statistical population, due to constraints (recording period, cost, etc.)
- A researcher would instead observe a **statistical sample** from the population. He will estimate an econometric model to understand the **properties on the population as a whole**.

Data Generating Process

Definition: Data Generating Process

A **Data Generating Process (DGP)** is a process in the real world that "generates" the data (or the sample) of interest

Example: Data Generating Process

Let us assume that there is a linear relationship between interest rates in two countries (R, R^*), their forward (F) and their spot exchange rate (S).

$$\frac{F}{S} = \frac{1+R}{1+R^*}$$

This non-arbitrage relationship (CIP) can be used in the foreign exchange market to determine the forward exchange rate

$$\mathbb{E}[F|S=s, R=r, R^*=r^*] = s * \frac{1+r}{1+r^*}$$

This relationship is the **Data Generating Process** for F

The equivalent of population for time series econometrics is the DGP.
NB: note that I use R to describe the random variable and r to describe its realization

Econometrics Challenge

The challenge of econometrics is to draw conclusions about a DGP (or population), after observing only one realization $\{x_1, \ldots X_N\}$ of a random sample (the dataset).

Data Types

In econometrics, sets can be mainly distinguished in three types:

1. Cross-sectional data
2. Time series data
3. Panel data

Cross-Sectional Data

Cross-sectional data are the most common type of data encountered in statistics and econometrics.

- Data at the entities level: banks, countries, individuals, households, etc.
- **No time dimension**: only one "wave" or multiple waves of different entities
- Order of data does not matter: no time structure

Time Series Data

Time series data are very common in financial econometrics and central banking. They entail specific estimation methods to do the **time-dependence**.

- Data for a single entity (person, bank, country, etc.) collected at multiple time periods. Repeated observations of the same variables (interest rate, GDP, prices, etc.)

- Order of data is important!

- The observations are typically not independent over time

- In this case, the notion of population corresponds to the **Data Generating Process (DGP)**

Panel Data

Also called longitudinal data. They contain the most information and allow for more complex estimation and analysis.

- Data for multiple entities (individuals, firms, countries, banks, etc.) in which outcomes and characteristics of each entity are observed at multiple points in time
- Combine cross-sectional and time-series information
- Present several advantages with respect to cross-sectional and time series data, depending on the topic at hands

Econometric Model

Definition: Econometric Model

An econometric model specificies the statistical relationship between different economic variables, that are expected to be stable over time

1. **Parametric model:** fully characterization of the relationship by a **set of parameters** θ and a **link function** f supposed to be known; the specification can be linear or non linear, and includes some randomness ϵ

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

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

Empirical Strategy

The general approach of (financial) econometrics is as follows:

1. Specification of the model
2. Estimation of the parameters
3. Diagnostic tests
   - Significance tests
   - Specification tests
   - Backtesting tests
   - etc.
4. Interpretation and use of the model (forecasting, historical studies, etc.)

Refresher: Random Variables

- Mathematicians are formalizing and modeling randomness via the concept of **random variables**.
- Pay attention: a random variable is neither random (it is formalized via laws and distributions), not a variable (it is a function)
- A **random variable** is a function $f : \Omega \mapsto \mathcal{R}$ that assigns to a set of outcome Ω a **value**, often a real number.
- The probability of an outcome is equal to its **measure** divided by the measure of all possible outcomes
  - Example: obtaining an even number by rolling a dice: $\{2, 4, 6\}$
  - Probability to obtain an even number by rolling a dice: $m(\{2, 4, 6\})/m(\{1, 2, 3, 4, 5, 6\}) = \frac{1}{2}$ *(here, the measure simply "counts" the outcomes with equal weights)*

Random Variables (II)

- Random variables are the "building block" of statistics:
  - Random variables are characterized by their distribution (generating function, moments, quantiles, etc.)
  - The behavior of two or more random variables can be characterized by their dependence/independence, matrix of variance-covariance, joint distribution, etc.
  - The main theorems in statistics (law of large numbers, central limit theorem, etc.) leverage the properties of random variables

Refresher: Probability Distributions for Continuous Variables

Definition: Probability Distribution

The probability distribution of a random variable describes how the probabilities of the outcomes are distributed. How more likely is one outcome compared to another? Is there more upside risk or downside risk? Etc.

Distributions are equivalently represented by their:

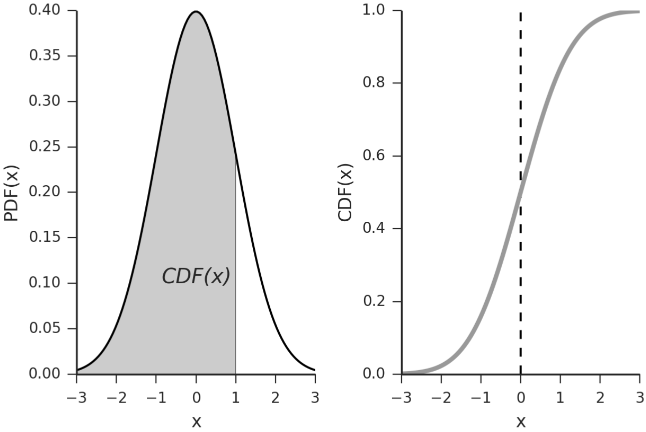
1. **The probability density function (pdf)**
2. **The cumulative distribution function (cdf)**

The pdf is also necessary to define distributions' moments (see after).

- **The probability density function (pdf)** usually denoted $f_X(x0)$ (remember than X stands for the random variable, while $x0$ stands for a realization of the random variable) represents the relatively likelihood that the random variable X will fall within a small neighborhood of $x0$ (infinitesimal concept). It is easier to conceptualize the pdf via the cdf
- **The cumulative distribution function (cdf)** usually denoted $F_X(x0)$ represents the probability that the random variable will be lower than $x0$. It cumulates the pdf ("all the small neighborhoods") such that:

$$F_X(x0) \equiv \mathbb{P}[X \leqslant x0] = \int_{-infty}^{x0} f_X(h)dh$$

Link between PDF and CDF



Moments: Overview

	Formula	Interpretation
Mean	$\mathbb{E}[X_t] = \mu$	Central tendency
Variance	$\mathbb{V}[X] = \mathbb{E}\left[(X_t - \mu)^2\right] = \sigma^2$	Dispersion around μ
Skewness	$\mathbb{S}[X] = \mathbb{E}\left[(X_t - \mu)^3\right] = \text{sk}$	Symmetry
Kurtosis	$\mathbb{K}[X] = \mathbb{E}\left[(X_t - \mu)^4\right] = \kappa$	Tail heaviness

Moments: In Practice

- The moments allow to characterize the shape of the returns distribution
- However, the theoretical moments are **unobservable** and need to be estimated
- Assume that we have a sample $\{x_1, \ldots, x_T\}$ realizations of the sequence of X_t

Estimator

Definition: Estimator

An **estimator** is any function $F(x1, \ldots, x_t)$ of a sample. Note that any descriptive statistics is an estimator (a simple one)

Example: Sample Mean

The sample mean (or overage) of a sample is an estimator of the (theoretical) mean $\mathbb{E}[X_t] = \mu$.
The estimator is simply: $\hat{\mu_t} \equiv \bar{X}_t = \frac{1}{T} \sum_{t=1}^{T} x_t$

Example: Variance

Example: Sample Mean

Assume that the observations are drawn from *i.i.d* random variables.
The **sample variance** $si\hat{g}ma_T^2 = \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \bar{x}_t)^2$

Note: the denominator is equal to *T-1* as to define a sample variance corrected for the small sample bias.

Sampling Distribution

Fact

An estimator $\hat{\theta}$ is a **random variable**

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

Definition: Sampling Distributions

The probability distribution of an estimator is called the **sampling distribution**

The sampling distribution is described by its moments, such as expectations, variance, skewness, etc.

Point Estimate

Definition: Estimate

An estimate is the realized value of an estimator (e.g. a number, in a case of a point estimate) that is obtained for a particular value $x0$. Often noted as $\hat{\theta}(x0)$ for the estimator $\hat{\theta}$

Estimate of a linear regression

- We are interested in the following DGP: $Y = \alpha + \beta * X + \epsilon$, where we observe a joint sample $y1, \ldots, y_T, x1, \ldots, x_T$.
- We have an estimator (for instance an OLS) of $\hat{\alpha}, \hat{\beta}$
- Then, for any value of $X = x_0$, we can simply project the **conditional expected estimate** $y_0 = \hat{\alpha} + \hat{\beta} * x_o + \hat{\epsilon}$
- If the estimator is unbiased, the fitted residuals $\hat{\epsilon} = y_t - \hat{\alpha} - \hat{\beta} * x_t$ are centered on **average**:$\mathbb{E}\left[\hat{\epsilon}\right] = 0$. This is why the residual disappear from the estimate of the conditional expected estimate in an OLS... but no bias doesn't mean no variance !
- ϵ and consequently $\hat{\epsilon}$ are random variables: their distribution will determine the distribution of the estimator $\hat{\theta} = \hat{\alpha}, \hat{\beta}$

What Constitutes a Good Estimator?

The idea is to study the properties of the **sampling distribution** of the estimator $\hat{\theta}$, and especially its moments, such as:

- $\mathbb{E}[\hat{\theta}]$ for the biais
- $\mathbb{V}[\hat{\theta}]$ for the precision
- $\mathbb{S}[\hat{\theta}]$ for the symmetry
- $\mathbb{K}[\hat{\theta}]$ for the tail-risks
- etc.

Estimators Properties

Estimators are compared on the basis of a variety of attributes

1. **Finite sample properties** (or finite sample distributions) investigate how the estimator behave when the observational sample is relatively limited (a few hundreds to a few thousands of observations max)

2. However, these properties rely on a distributional assumption (usually normality or gaussianity), that may be difficult to test.

3. When the normality assumption is no longer valid (and the finite sample distribution is unknown), estimators are evaluated on the basis on their **large sample**, or **asymptotic properties**

Finite Sample Theorem

Theorem: Finite Sample Distributions

If we assume that, with a sample of size T, generated from a stochastic process with i.i.d random variables $X_1, X_2, \ldots, X_T$ with $X_t \sim \mathcal{N}(\mu, \sigma^2)$, then the estimators of the sample mean $\hat{\mu_T}$ and the estimator of the sample variance/population variance $(T-1)\frac{\hat{\sigma_T^2}}{\sigma^2}$ have a **finite sample distribution**

$$\hat{\mu_T} = \mathcal{N}\left(\mu, \frac{\sigma^2}{T}\right) \qquad \forall\ T\ \in\ \mathbb{N}$$

$$\frac{T-1}{\sigma^2}\hat{\sigma^2} \sim \chi^2(T-1) \qquad \forall T\ \geqslant\ 2$$

Example: Finite Sample Distribution

Under the Gaussianity assumption, with a sample size of $T = 10$, then:

$$\hat{\mu} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{10}\right) \qquad 9 \times \frac{s\hat{i}gma_T^2}{\sigma^2} \sim \chi^2(9)$$

Difficulties with Finite Sample Inference

In most cases, it is impossible to derive the **exact/finite sample distribution** for the estimator (or any transformation of the estimator).

1. In some cases, the exact distribution of $\{X1, X2, \ldots, X_T\}$ is known, but the estimator function $S()$ (also called the "statistics") is too complicated

$$\hat{\theta} = S(X_1, X_2, \ldots, X_T) \sim ??? \qquad \forall\ T \ \in \mathbb{N}$$

2. In most cases, the distribution of the stochastic process $\{X1, X2, \ldots, X_T\}$ is unknown

$$\hat{\theta} = S(X_1, X_2, \ldots, X_T) \sim ??? \qquad \forall\ T \ \in \mathbb{N}$$

Asymptotic Properties

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 sufficiently large (usually, more than 10k)

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

Convergence in Probability

There are many types of convergence, but typically applied statisticians are interested in:

1. Convergence in probability
2. Convergence in distribution

- **Convergence in probability** (also called "mean squared convergence" or "almost sure convergence")

$$\hat{\theta_T} \xrightarrow{p} \theta \qquad \text{for } \theta \in \mathbb{R}$$

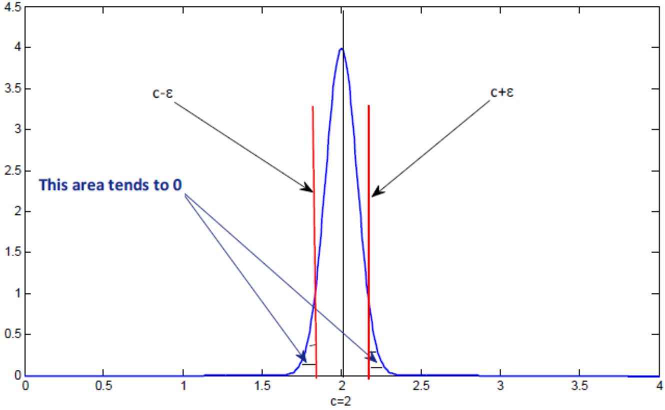
- Technically, for a stochastic process $\theta_T{}_{-\infty}^{+\infty}$

$$\hat{\theta_T} \xrightarrow{p} \theta \quad \Leftrightarrow \quad \lim_{T \to +\infty} \mathbb{P}\left[|\theta_T - \theta| > \epsilon\right] = 0$$

- The convergence in probability is used to derive the **consistency property** of estimators

Illustration: Distribution in Probability

$$\hat{X_T} \xrightarrow{p} c \text{ if } \Leftrightarrow \quad \lim_{T\to+\infty}\mathbb{P}\left[|X_T - c| > \epsilon\right] = 0$$



Interpretation

1. The distribution of the sample moments is highly concentrated around the true value (unknown) of the population moment when the sample size T tends to infinity

2. In other words, over a very large sample of data (for instance, hundred of thousands of observations) the **moments estimated in the sample** are then very "close" to the true value of **the moments in the population**

Convergence in Distribution

- **Convergence in distribution** is interested in the distribution of the bias (the distance between the estimator and the true value)

$$\sqrt{T}\left(\hat{\theta_T} - \theta_0\right) \sim \mathcal{N}(\mu_., \sigma_.) \qquad \mathcal{N} \text{ is the most common}$$

- Technically, for a stochastic process ${X_T}_{-\infty}^{+\infty}$, with a cdf $F_T(.)$; X_T is said to **converge in distribution** to a random variable X if:

$$X_t \xrightarrow{d} X \quad \Leftrightarrow \quad F_T(x) = F(x) \qquad \forall\ x\ \in\ \mathbb{R}$$

- The convergence in distribution is used to derive the **asymptotic distribution** of the estimators and to make **inference** (tests) about the true value of the parameters

Stylized Properties

Fan and Yao (2015, the Elements of Financial Econometrics) identify 8 main "stylized facts"

1. Stationarity
2. Absence of autocorrelations
3. Heavy tails
4. Asymmetry
5. Volatility clustering
6. Aggregational Gaussianity
7. Long range dependence
8. Leverage effect

Stationarity

Example

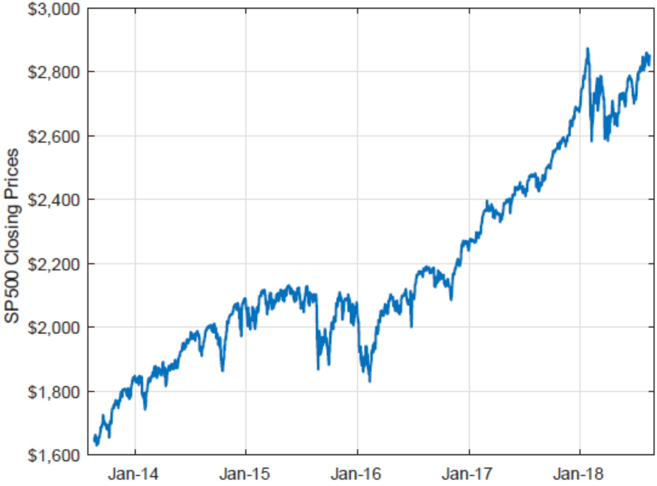
In general, prices are non-stationary but returns are stationary

Definition: Weak Stationarity (Second Order Stationarity)

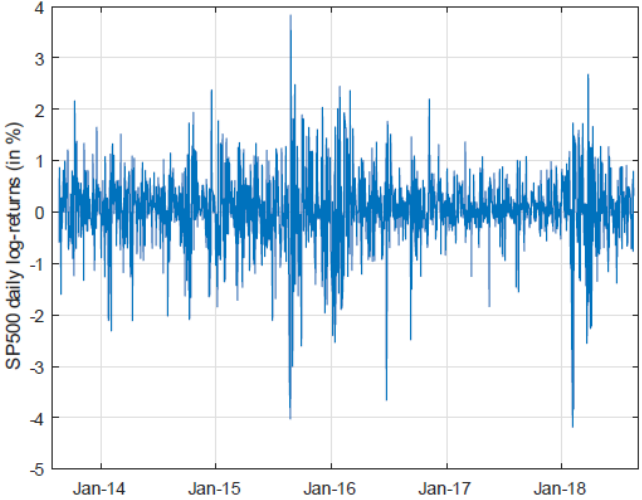
A stochastic process $X_{t t\in\mathbb{Z}}$ is weakly stationary if and only if:

- $\mathbb{E}(X_t^2) \ < \ \infty \quad \forall \ t \ \in \ \mathbb{Z}$
- $\mathbb{E}(X_t) = \mu \quad \forall \ t \ \in \ \mathbb{Z}$ doesn't depend on t
- $\mathrm{Cov}(x_t, x_{t+h}) \ = \ \mathbb{E}[(x_{t+h} - m)(x_t - m)] \ = \ \gamma_h \quad \forall \ (t, h) \ \in \ \mathbb{Z}^2$ desn't depend on t

Prices are Usually Non-Stationary...



… But Returns Are !



Absence of autocorrelations

Fact: Common absence of autocorrelations

The autocorrelations of assets returns (in general) are often insignificant, except for very small intraday time scales (around 20 minutes) for which the microstructure effects come into play

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

Definition: Autocorrelation

The autocorrelation, denoted $\rho(k)$ of a weak stationary process X_t is the correlation between the values of the process at different times:

$$\rho_k \;=\; \mathrm{Corr}(X_t, X_{t-k}) \;=\; \frac{\mathbb{E}\left[(X_t-\mu)(X_{t-k}-\mu)\right]}{\mathbb{V}(X_t)} = \frac{\gamma_k}{\sigma^2}$$

with $\mu = \mathbb{E}[X_t]$, $\sigma^2 = \mathbb{V}(X_t),\ \forall\ t$ and γ_k the autocovariance of order k

Measuring Sample Autocorrelation

Definition: Sample Autocorrelation

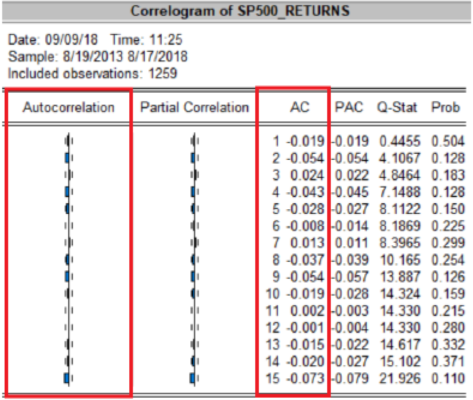
The **sample autocorrelation**, denoted $\hat{\rho}(k)$ of a weak stationary process X is an estimator of $\rho(k)$ defined as:

$$\hat{\rho}_k = \text{corr}(X_t, X_{t-k}) = \frac{1}{(T-k)\hat{\sigma^2}} \sum_{t=k+1}^{R} (X_t - \hat{mu})(X_{t-k} - \hat{mu})$$

where $\hat{\sigma^2}$ and $\hat{\mu}$ are consistent estimators of $\mu = \mathbb{E}(X_t)$ and $\sigma^2 = \mathbb{V}(X_t) \qquad \forall t$

Example: Autocorrelation of the SP500

The **Autocorrelation Function (ACF)** (or **correlogram**) represents the sample autocorrelation for different lags, from $k = 1$ to a maximum lag order, for instance $k = 15$



Correlogram of SP500_RETURNS

Date: 09/09/18 Time: 11:25
Sample: 8/19/2013 8/17/2018
Included observations: 1259

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.019	-0.019	0.4455	0.504
		2	-0.054	-0.054	4.1067	0.128
		3	0.024	0.022	4.8464	0.183
		4	-0.043	-0.045	7.1488	0.128
		5	-0.028	-0.027	8.1122	0.150
		6	-0.008	-0.014	8.1869	0.225
		7	0.013	0.011	8.3965	0.299
		8	-0.037	-0.039	10.165	0.254
		9	-0.054	-0.057	13.887	0.126
		10	-0.019	-0.028	14.324	0.159
		11	0.002	-0.003	14.330	0.215
		12	-0.001	-0.004	14.330	0.280
		13	-0.015	-0.022	14.617	0.332
		14	-0.020	-0.027	15.102	0.371
		15	-0.073	-0.079	21.926	0.110

Asymmetry

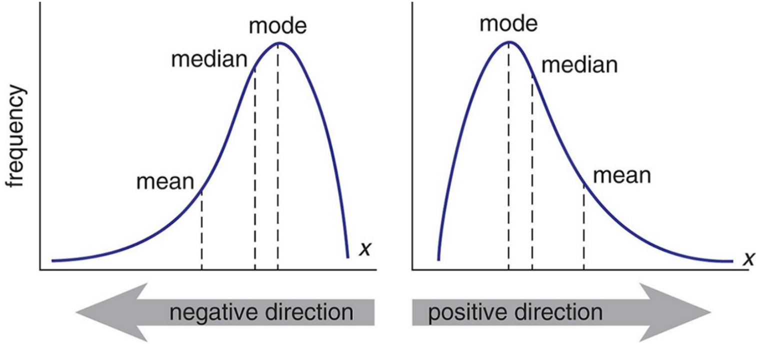
Stylized Fact: Heavy Tails

The distribution of many financial variables, including asset returns, are often **asymmetric** and **negatively skewed**

- Asymmetry is defined by the skewness, which is the third-order moment (see before)
- This reflects the fact that the downturns of financial markets are often much steeper than the recoveries
- Investors tend to react more strongly to negative news that to positive news

Skewness

There are different shapes of kurtosis



Heavy Tails

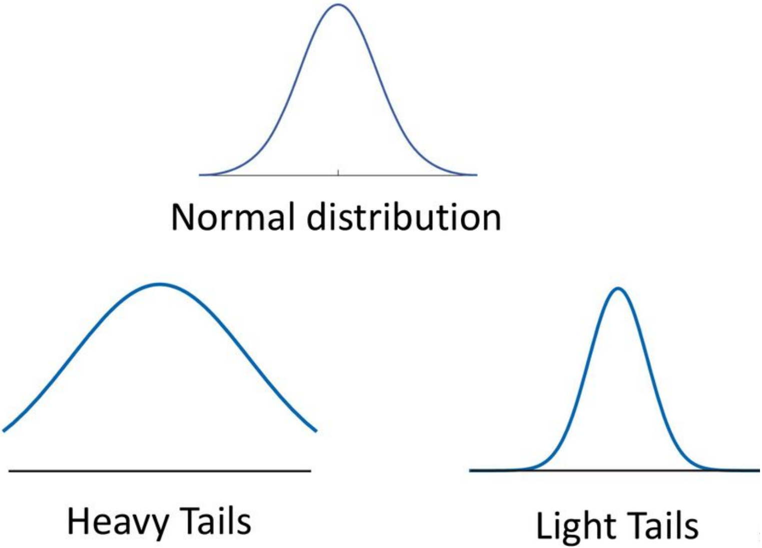
Fact: Heavy Tails

The probability distribution of many financial variables, including asset returns, often exhibit **heavier tails** than those of a normal distribution

- "Heavier tails" are rigorously defined by the kurtosis, which is the fourth-order moment (see before)
- Mandelbrot (1963) recognized the heavy-tailed, highly peaked nature of certain financial time series
- These heavy tails can be explained by risk aversion, heard behavior, market microstructure (illiquidity, asymmetric information, etc.)

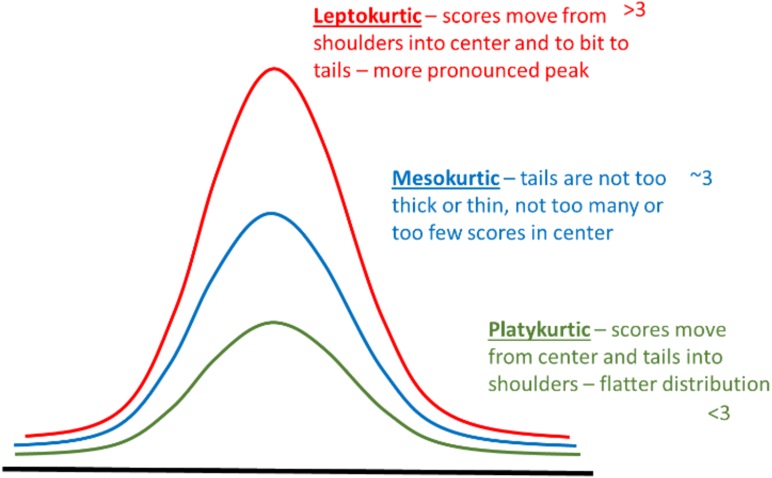
Normal versus Tails

There are different shapes of kurtosis



Forms of Kurtosis (Fat Tails)

There are different shapes of kurtosis:



Volatility Clustering

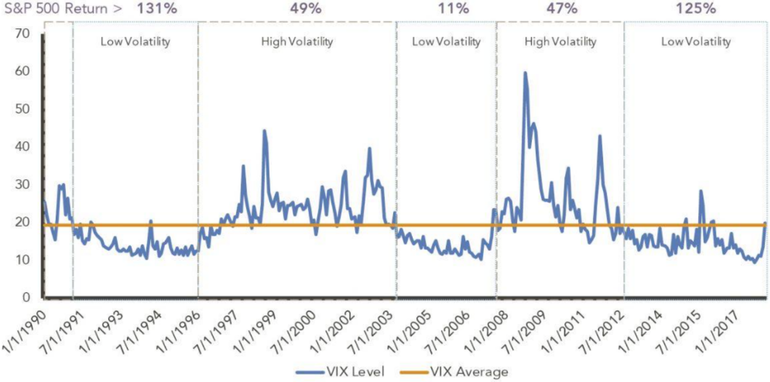
Definition: Volatility Clustering

- **Volatility clustering** means that large price changes (i.e. returns with large absolute values or large squares) occur in clusters
- Large price changes tend to be followed by large price changes (up and down)
- Periods of tranquility alternate with periods of high volatility (volatility regimes)

Note: volatility clustering is the consequence of the autocorrelation of the squared returns

Volatility Regimes: US VIX

The VIX is the implied volatility fo the US SP 500:



Aggregational Gaussianity

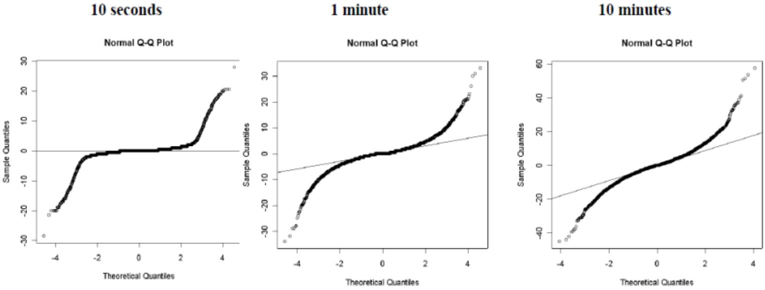
Definition: Aggregational Gaussianity

- Asset returns over k days is simply the aggregation of k daily returns
- When the time horizon k increases, the central limit theory says that the distribution of returns over a long-time horizon (a few months) tends toward a **normal distribution**

- Aggregational gaussianity implies that over long horizons, the pecularities of financial time series over short-term horizon (skewness, kurtosis, etc.) tend to vanish
- However, in finance, people are mostly interested in relatively short-term movements, suggesting that working under the gaussianity assumption is often not appropriate

Aggregational Gaussanity In Practice

The VIX is the implied volatility fo the US SP 500:



Long Range Dependence

Definition

- At the difference of log returns or standard returns, daily squared returns and absolute returns often exhibit significant autocorrelations
- Those autocorrelations are persistent, indicating possible **long-memory** properties

Those autocorrelations become weaker and less persistent when the sampling interval is increased to a week or a month

Long Range Dependence

SP 500 Returns (left) and squared returns (right)

Figure: ACF for the S&P500 returns

Date: 09/09/18 Time: 15:19
Sample: 8/19/2013 8/17/2018
Included observations: 1259

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.019	-0.019	0.4455	0.504
		2	-0.054	-0.054	4.1067	0.128
		3	0.024	0.022	4.8464	0.183
		4	-0.043	-0.045	7.1488	0.128
		5	-0.028	-0.027	8.1122	0.150
		6	-0.008	-0.014	8.1869	0.225
		7	0.013	0.011	8.3965	0.299
		8	-0.037	-0.039	10.165	0.254
		9	-0.054	-0.057	13.887	0.126
		10	-0.019	-0.028	14.324	0.159
		11	0.002	-0.003	14.330	0.215
		12	-0.001	-0.004	14.330	0.280
		13	-0.015	-0.022	14.617	0.332
		14	-0.020	-0.027	15.102	0.371
		15	-0.073	-0.079	21.926	0.110

Figure: ACF for the S&P500 squared returns

Date: 09/09/18 Time: 11:25
Sample: 8/19/2013 8/17/2018
Included observations: 1259

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.302	0.302	114.94	0.000
		2	0.251	0.176	194.81	0.000
		3	0.288	0.196	299.82	0.000
		4	0.251	0.116	379.50	0.000
		5	0.115	-0.052	396.30	0.000
		6	0.135	0.018	419.45	0.000
		7	0.095	-0.020	430.83	0.000
		8	0.108	0.044	445.63	0.000
		9	0.102	0.043	458.93	0.000
		10	0.123	0.060	478.14	0.000
		11	0.084	0.003	487.21	0.000
		12	0.050	-0.041	490.35	0.000
		13	0.048	-0.020	493.24	0.000
		14	0.040	-0.012	495.29	0.000
		15	0.031	0.010	496.51	0.000

The ARCH effect

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

ARCH effect is important in finance, because it describes patterns on the dynamic of financial volatility

Leverage Effect

Definition: the Leverage Effect

- Assets returns are negatively correlated with the changes in their volatilities
- This negative correlation Corr (returns, vol) is called the **leverage effect**

Financial explanations

- An asset price declines, companies mechanically become more leveraged (debt-to-equity ratio up) and riskier: therefore, 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 prices appreciation due to declined volatilities