Empirical Methods in Finance (MFE230E)
Week 1: Course introduction, Introduction to time series models

Martin Lettau

Spring 2019

Haas School of Business

Organization

Instructor: Professor Martin Lettau, F608, 3-6349

Office Hours: W 11:00 – 12:00

GSI: Paulo Manoel and Ruochen Zeng (`mfe230e@gmail.com`)

GSI for lab session: Chris Jauregui (`mfe230e.python@gmail.com`)

Class Time: Mo 9:00–12:00/1:30–4:30 (F320) and Wed 6:00–9:00 (N100)

Lab Time: Fr 10:00–12:00 in C230

Discussion sections: Tu 2:00–4:00 in C230

GSI office hours: TBA

Course website: `https://bcourses.berkeley.edu/courses/1478108`

Exams: Midterm: April 17, 6:00–7:30, Final: May 22, 9:00–12:00

Please read the course syllabus carefully!

Communication:

- **Please use the e-mail address** `mfe230e@gmail.com` **for all communication!**
- For all python-related questions, use `mfe230e.python@gmail.com`
- Do not use the bCourses e-mail client!
- I encourage the use of the bCourses "Discussion" module to communicate with each other

Schedule

- Two 3-hour lectures per week.
- In addition to the lectures, there will be a weekly lab session in which we will implement the theoretical material covered in class and use empirical examples to illustrate empirical methods. The lab sessions will use Python as programming language and Python is the required language for this class.
- The weekly GSI discussion session will focus on 'problem set-style' examples.

Lecture Notes and other Course Material

Lecture notes will be available on bCourses under the study.net tab a few days prior to class. Please bring the lecture notes to each class. All additional reading material will be announced before class and will also be available on bCourses.

Grading

- Final exam: 50%
- max(Midterm, Final): 20%
- Problem sets: 10%
- Class participation: 20%

Exams

The exams are closed-book but you can bring one double-sided cheat sheet. You may also use a financial calculator during the exams.

You may not refer to any materials during the exams, including other textbooks, your homework or solutions from this class, the manual for your calculator, etc. You may not bring a laptop, PDA, cell phone, or any other device that allows you to access the internet. All examinations will be audio and videotaped for the purpose of assuring academic integrity.

Honor Code

All students agree to abide by the Berkeley Campus Code of Student Conduct, http://students.berkeley.edu/uga/cs/general/policy.htm, and by the terms of the the supplemental MFE Code of Student Conduct, which you signed upon entering the program.

Students are not allowed to use problem set solutions from previous classes in any way! Only original material will be accepted. Solutions with screen shots are not accepted. Students must submit original documents if requested.

Problem Sets

There will be weekly problem sets. These problem sets are a crucial component of the course as you will apply what you have learned in the lectures. The homework sets will mostly consist of empirical exercises using real financial data. You will have to create your own library of Python code that will be helpful throughout your career in finance. Problem sets may be completed in group work. Only a single group solution should be handed in (with names of all group members). However, you should **not** split the assignments among group members!

Problem sets have to be submitted via bCourses before the deadline stated on the problem set. Late problem sets will not be accepted. No exceptions and no individual extensions! If for whatever reason you cannot log into bCourses to submit a problem set, you need to e-mail the problem set to `mfe230e@gmail.com` before the deadline to receive credit.

Code for problem sets

The general rule for code for problem sets is that you may use *basic* Python commands but not full built-in routines.

Example: The problem set asks you to estimate a GARCH model by maximizing the likelihood function. You may NOT use the Python `arch` package:

```
from arch import arch_model
am = arch_model(returns)
res = am.fit(update_freq=5)
```

Instead, you have to solve for the likelihood function first and then use a numerical optimization package (e.g. `scipy.optimize`). If you are in doubt about what is acceptable, please ask!

Class Participation

Class participation is crucial to the success of this class. You will get the most out of this course if you come to each lecture prepared and ask questions. We will cover a lot of material in a short amount of time and many of the key concepts are challenging. I urge you to ask clarifying questions or to ask me to explain a concept again. If, for some reason, you have to miss a class, please get the notes from one of your classmates and review it carefully. Class participation will count 10% towards your final grade.

To ensure a broad class participation, I will follow a “30-second rule” to give students time to think about questions before calling on people. It is important that every student participates actively in the class. I will cold-call on individual students!

Ethics and Etiquette

I expect students to refrain from behavior that has been demonstrated to interfere with a positive classroom experience.

- Come to class prepared.
- **PLEASE arrive on time.**
- Bring your name cards.
- Do not use phones or laptops to surf the Web, check e-mail, etc.
- Put away reading material that is not course-related.
- Do not talk among yourselves in class. If you have a question, please raise your hand.

Honor Code

All students agree to abide by the Berkeley Campus Code of Student Conduct, `http://students.berkeley.edu/uga/cs/general/policy.htm`, and by the terms of the the supplemental MFE Code of Student Conduct, which you signed upon entering the program.

In particular, students are not allowed to use problem set solutions from previous classes in any way!

Only original material will be accepted. Solutions with screen shots are not accepted. Students must submit original documents if requested.

Course Objectives

- Understand theory and use of financial time series in depth.
- Develop a toolbox of econometric methods.
- Learn how to apply these methods to financial data.
- Learn how to test asset pricing models.
- Understand which models work well and which ones do not.
- Understand *why* models work well or don't work well (remember: *every* model is wrong!).
- Understand appropriate limits of econometric methods in finance.

Required: David Ruppert and David Matteson, *Statistics and Data Analysis for Financial Engineering with R Examples, 2nd edition*, Springer, 2015

Optional: (but very useful) material:

- James Hamilton, *Time Series Analysis*, Princeton University Press, 1994
  An invaluable reference textbook for time series methods
- John Cochrane, *Time Series for Macroeconomics and Finance*, notes available at `http://faculty.chicagobooth.edu/john.cochrane/research/papers/time_series_book.pdf`. Very useful summary of time series tools used in finance
- Ruey Tsay, *Analysis of Financial Time Series*, 3rd ed., Wiley, 2010
  Strongly recommended but somewhat more advanced than Ruppert
- Campbell, Lo, MacKinlay, *The Econometric of Financial Markets*, Princeton, 1997
- John Cochrane, *Asset Pricing*, revised edition, Princeton, 2001
  Very good treatment of GMM and cross-sectional asset pricing tests
- Shumway and Stoffer (SS): *Time Series Analysis and Its Applications, EZ Edition*, pdf available on study.net (under CC BY-NC 4.0 license).
  Useful background text for time series

Reference: Every quantitative finance practitioner should have a good reference statistics textbook, e.g. DeGroot and Schervish *Probability and Statistics* as well as a comprehensive econometrics text (e.g. Hayashi's *Econometrics*) on their bookshelf.

The 2013 Nobel Prize in Economics was awarded to three finance researchers: Gene Fama, Lars Hansen and Bob Shiller. The research of all three laureates will be featured in this class. I encourage you to watch their Nobel lectures (https://www.youtube.com/watch?v=WzxZGvrpFu4).

COURSE OUTLINE I

1. **Stochastic processes, linear time series, stationarity**
2. **Estimation of linear times series models**: Review of classical regressions model, large sample distribution of least squares estimator
3. **Estimation of non-linear time series models**: Maximum likelihood estimation, General Method of Moments
4. **Cointegration**: Present value relationships, tests of stock return predictability, analysis of the price-dividend ratio, long-horizon regressions

Course outline II

5. **The cross-section of stock returns**: CAPM, tests of the CAPM, empirical evidence, factor models, Fama-French models
6. **Models of time-varying volatility**: GARCH, stochastic volatility, MLE estimation
7. **Principal component models**: Applications to asset returns
8. **Kalman filter, state-space models**
9. **Bayesian models**: MCMC methods, bootstrapping, resampling (if time permits)

Prerequisite: MFE Statistics pre-program class I

Topic 1: Probability

1. Definitions, independence, Bayes' theorem (DGS chs. 1, 2)
2. Random variables (DGS chs. 3): Univariate and multivariate distributions, marginal and conditional distributions, important distributions
3. Moments (DGS ch. 4): Expectations, variance, moment generating function
4. Convergence of RVs (DGS ch. 6): Law of large numbers, central limit theorem
5. Markov chains (DGS ch. 3.10)

Topic 2: Statistics

1. Estimation (DGS ch. 7): Maximum likelihood, method of moments
2. Bayesian estimation (DGS ch. 7)
3. Sampling distributions of estimators (DGS ch. 8): χ^2, t, F distributions, limiting distributions, confidence intervals
4. Hypothesis testing (DGS ch. 9): Simple tests, power vs. size
5. Bayesian estimation: Priors and posterior, Bayesian inference

Prerequisite: MFE Statistics pre-program class II

Topic 3: Linear statistical model

1. Conditional expectation functions and projections (lecture notes): Conditional expectation functions, conditional variance, best linear predictor
2. Linear model (DGS ch. 11): Regression, small sample distribution, Gauss-Markov theorem

Topic 4: Introduction to linear time series methods

1. Introduction to linear time series (SS ch. 1 and 3)
2. Autocorrelation and partial autocorrelation functions
3. Dependence and stationarity
4. AR models

FOR STUDENTS WITH WEAKER STATISTICS BACKGROUND

If you have a weaker statistics background, **please let us know!**

If the pre-program statistics class was very difficult for you to follow, **please let us know!**

If the rest of the lecture today is challenging for you, **please let us know!**

Do NOT wait until the midterm!

Hang in there!

- **My commitment**: Make this class the most useful MFE class in the short-term (e.g. for interviews) and long-term (give you tools that you will use the rest of your career) ... but ...

- MFE230E is one one of the hardest and most intensive MFE classes!
- The first 3-4 weeks are the most difficult part of the class
- Some of you will struggle with the speed and workload at the beginning of the class!
- In the past, midterm evaluations are full of complaints
- **Hang in there!** The material will all make sense as we progress
- **Come to office hours! Work with your teammates! Ask questions! Do not hesitate to contact us!**

So, let's get started ...

- I have asked MFE recruiters what they expect candidates to know and what kind of questions they ask in interviews
- At the beginning of each Monday class, I will ask such "interview questions" about the material from the previous week
- I will NOT post the questions before class!
- So, let's start with "interview questions" about what you learned in the pre-program Stats class ..

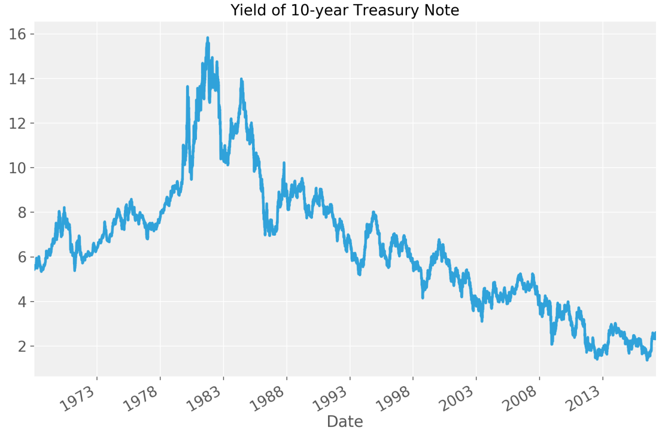
1. Introduction to time series models

    Examples

    ARMA Models

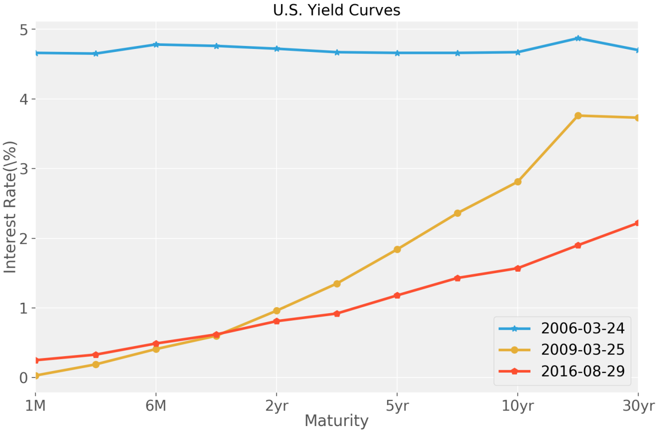
2. Stationarity

1. Time-series methods
2. Cross-sectional methods
3. Panel data methods: TS + CS



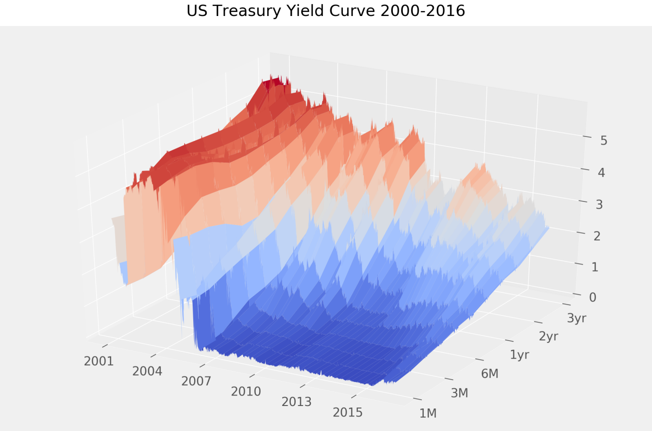

```
start = datetime.datetime(1968, 1, 1)
end = datetime.datetime(2017, 3, 26)
df = data.DataReader("^TNX", 'yahoo', start, end)
```

Cross section: Yield curve on a given day




```
df = pd.read_csv('https://raw.githubusercontent.com/mlettau/Data/master/MFE230E/
    Week-1/FRB_H15.csv', skiprows= 5)
```

Panel: Yield curve over time



- Linear time series processes: Univariate ARMA models
- Serial correlation and stationarity
- Non-stationarity and unit roots
- Cointegration
- Forecasting
- Multivariate ARMA processes
- Reading:
  - SS chapters 1 and 3 (background)
  - Ruppert chapter 9 (chapter 10 will not be covered but is optional reading)
  - Cochrane's *Time Series for Macroeconomics and Finance* notes
  - More advanced: Hamilton chapters 1, 2, 3, 4, 10, 11

Definition 1 (Time series).

A **time series** is a set of repeated observations over time of the same variable, such as GDP or a stock return. We can write a time series as

$$\{x_1, x_2, \ldots x_T\} \text{ or } \{x_t\}_{t=1}^{T}.$$

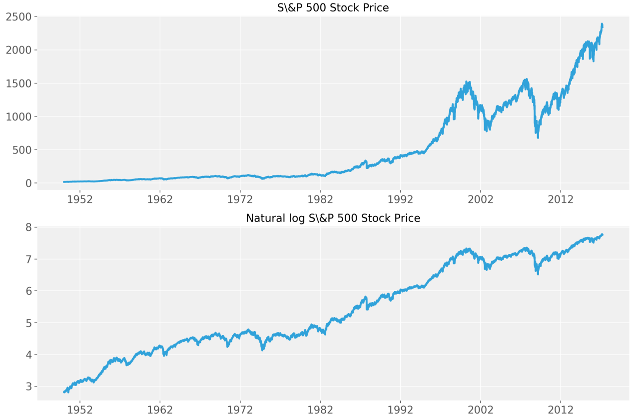
Each x_t is a **random variable**.

- Example of the joint distribution of $\{x_t\}_{t=1}^{T}$:

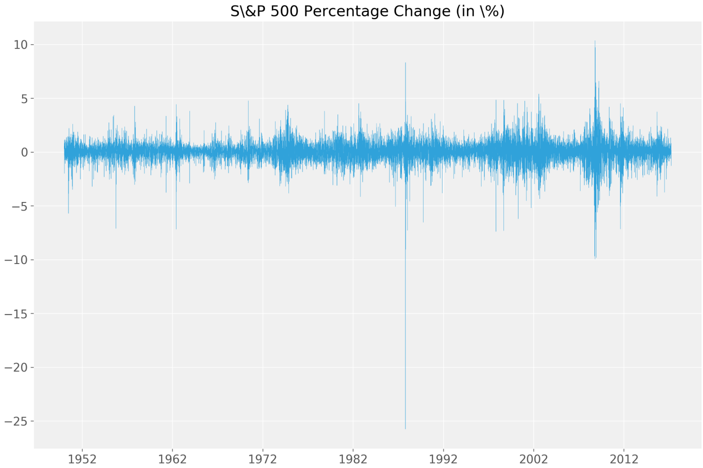
$$x_t = \epsilon_t, \qquad \epsilon_t \sim i.i.d.\ \mathcal{N}\left(0, \sigma_\epsilon^2\right),$$

- *i.i.d*: identically and independently distributed
- Some financial time series are (close to) i.i.d., but most are not

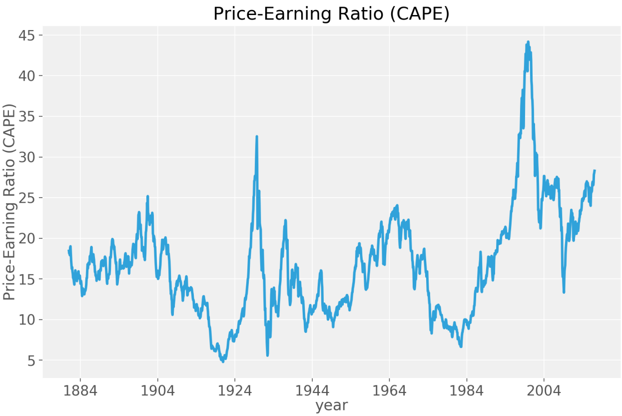
S&P 500 Stock Price




```
start = datetime.datetime(1953, 1, 1)
end = datetime.datetime(2017, 3, 26)
df = data.DataReader("^GSPC", 'yahoo', start, end)
```

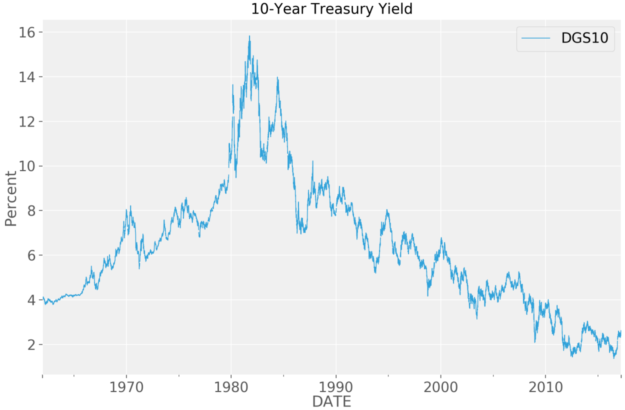

S\&P 500 Percentage Change (in \%)
10
5
0
−5
−10
−15
−20
−25
1952
1962
1972
1982
1992
2002
2012

SHILLER'S P/E RATIO: CAPE

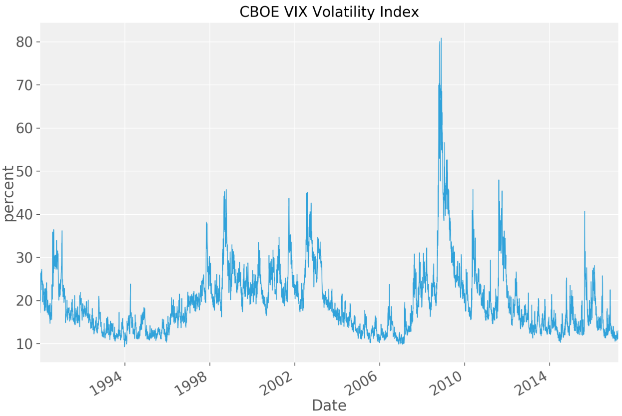



```
import quandl
df = quandl.get("MULTPL/SHILLER_PE_RATIO_MONTH")
```

10-YEAR TREASURY YIELDS




```
start = datetime.datetime(1962, 1, 1)
end = datetime.datetime(2017, 3, 26)
df = data.DataReader("DGS10", 'fred', start, end)
```




```
start = datetime.datetime(1990,1,1)
end = datetime.datetime(2017, 3, 26)
df = data.DataReader("VIX", 'yahoo', start, end)
```

Goal: Parsimonious but flexible models

- Note: Typically we have only one sample of the time series!
- Thus, we need a parsimonious class of models with few parameters to estimate the joint distribution of $\{x_t\}_{t=1}^{T}$
- Starting point: **Linear time series models**
- Quick review of material from week 10 of Stats Pro-program class

Definition 2 (White noise).

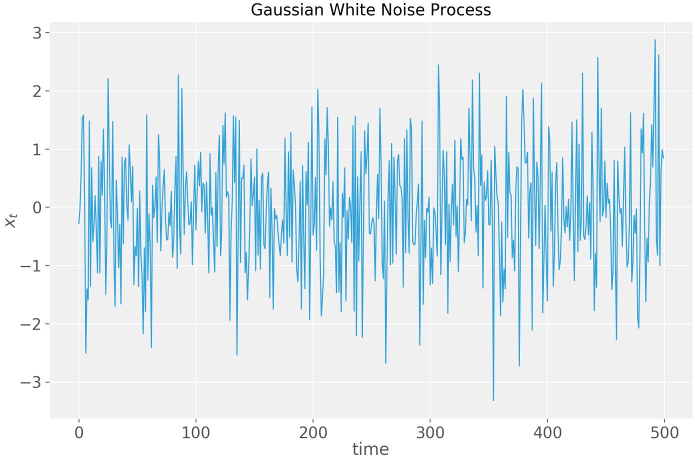
A process $\{\epsilon_t\}$ is called **white noise (WN)** if

1. $\mathrm{E}(\epsilon_t) = \mathrm{E}(\epsilon_t | \text{all information at } t-1) = 0.$
2. $\mathrm{E}(\epsilon_t \epsilon_{t-1}) = \mathrm{Cov}(\epsilon_t \epsilon_{t-1}) = 0$
3. $\mathrm{Var}(\epsilon_t) = \mathrm{Var}(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \ldots) = \mathrm{Var}(\epsilon_t | \text{all info at } t-1) = \sigma_\epsilon^2.$

If $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, then ϵ_t is called **Normal White Noise (NWN)**

- The second property is the absence of any **serial correlation or predictability**.
- The third property is **conditional homoskedasticity**, or a constant conditional variance.
- Question: Which financial time series shown so far might be well-modeled by a white noise process?

White Noise simulation



- Many financial TS exhibit **serial correlation**: Knowing today's stock price helps to forecast tomorrow's stock price
- Question: Which financial time series exhibit strong serial correlation?
- Definition:

$$\textbf{Autocovariance function: } \gamma_j \equiv \mathrm{Cov}(x_t, x_{t-j}) = \mathrm{E}[(x_t - \mu_x)(x_{t-j} - \mu_x)]$$

$$\textbf{Autocorrelation function: } \rho_j \equiv \frac{\gamma_j}{\gamma_0}$$

- The serial correlation of a time series will play an important role throughout the course, e.g. stationarity, unit roots, cointegration, long-run memory

Estimation: Method of Moment (sample equivalents)

$$\hat{\gamma}_j = \widehat{\mathrm{Cov}}(x_t, x_{t-j})$$

$$\hat{\rho}_j = \widehat{\mathrm{Corr}}(x_t, x_{t-j})$$

$$\sqrt{T}\,\hat{\rho}_j \to N(0, 1)$$

Question: What is the (asymptotic) standard deviation of $\hat{\rho}_j$ in a sample with T observations?

Testing whether a single autocorrelation coefficient is equal to zero:

$$\text{Under } H_0 : \rho_j = 0 : \sqrt{T}\hat{\rho}_j \rightarrow N(0,1)$$

Test whether the first h autocorrelation coefficients are jointly zero:

Box/Pierce-Ljung/Box test statistic

$$Q = T(T+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{T-k} \rightarrow \chi_h^2$$

- Consider a regression of x_t on x_{t-j}:

$$x_t = \phi_j x_{t-j} + \epsilon_t \tag{1}$$

- If $E[x_t] = 0$ and x_t is homoskedastic:

$$\widehat{\phi}_j = \frac{\widehat{\text{Cov}}(x_t, x_{t-j})}{\text{Var}(x_t)} = \widehat{\rho}_j$$

- Autocorrelation coefficients are regression coefficients in (1)
- ρ_j measures the unconditional correlation between x_t and x_{t-j}.

- Now consider the regressions

$$x_t = \phi_{1,1} x_{t-1} + \epsilon_t \tag{2}$$

$$x_t = \phi_{1,2} x_{t-1} + \phi_{2,2} x_{t-2} + \epsilon_t \tag{3}$$

- $\phi_{1,1}$ measures the unconditional correlation between x_t and x_{t-1}
- $\phi_{2,2}$ measures the correlation between x_t and x_{t-2} **net** of the correlation between x_t and x_{t-1}
- $\alpha_2 = \phi_{2,2}$ is called the **Partial Autocorrelation coefficient (PAC)**

- Intuition: The **partial autocorrelation coefficient** α_j measures the correlation of x_t and x_{t-j} **net of the autocorrelations** 1 **to** $j-1$.
- Consider the recursive system

$$x_t = \phi_{1,1} x_{t-1} + \epsilon_t$$

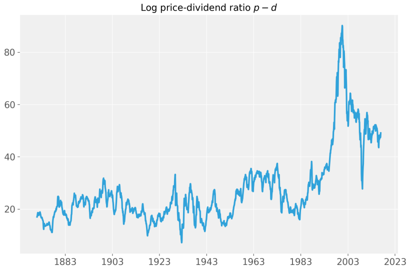
$$x_t = \phi_{1,2} x_{t-1} + \phi_{2,2} x_{t-2} + \epsilon_t$$

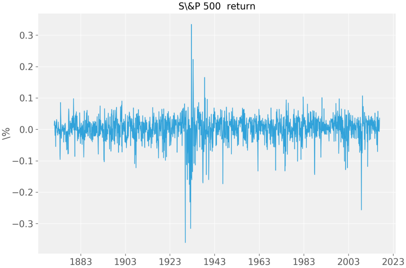
$$x_t = \phi_{1,j} x_{t-1} + \ldots + \phi_{j,j} x_{t-j} + \epsilon_t$$

- $\alpha_j = \phi_{j,j}$ is a measure of the marginal correlation of x_t and x_{t-j} net of the correlation between x_t and $x_{t+1}, x_{t+2}, \ldots, x_{t+h-1}$.

Definition 3.

$\alpha_j = \phi_{j,j}$ is called the **Partial Autocorrelation Function (PACF) of order** j.

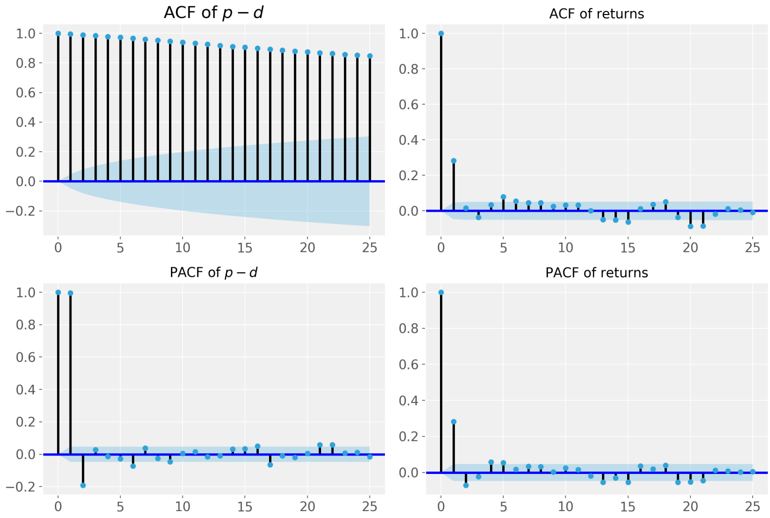





```
df = pd.read_excel("http://www.econ.yale.edu/~shiller/data/ie_data.xls"m
                                  sheetname = "Data", skiprows=7)
```

Sample: Jan 1881 - Mar 2017 = 1752 months or 146 years

Example: Log price-dividend ratio and S&P returns



lag	AC	PAC	Q-stat	Prob	AC	PAC	Q-stat	Prob
1	1.00	1.00	1739.63	0.00	0.28	0.28	140.20	0.00
2	0.99	-0.22	3460.56	0.00	0.02	-0.07	140.63	0.00
3	0.98	0.04	5162.73	0.00	-0.04	-0.02	142.91	0.00
4	0.98	-0.02	6846.06	0.00	0.03	0.06	144.98	0.00
5	0.97	-0.03	8509.93	0.00	0.08	0.06	155.81	0.00
6	0.97	-0.08	10152.34	0.00	0.06	0.02	161.30	0.00
7	0.96	0.05	11773.75	0.00	0.05	0.03	165.02	0.00
8	0.95	-0.03	13374.05	0.00	0.05	0.03	168.71	0.00
9	0.95	-0.05	14952.05	0.00	0.03	0.00	169.88	0.00
10	0.94	0.01	16507.61	0.00	0.03	0.02	171.72	0.00
11	0.93	0.02	18041.27	0.00	0.03	0.02	173.59	0.00
12	0.93	-0.02	19553.13	0.00	0.00	-0.02	173.59	0.00

Question: What is the standard deviation of $\hat{\rho}_1$ of $p - d$ and the S&P return?

Linear ARMA Processes

An important building block

The workhorse linear time series model is the **ARMA** class:

- **AR**: Auto-regressive
- **MA**: Moving-average

Building block: i.i.d. white noise process

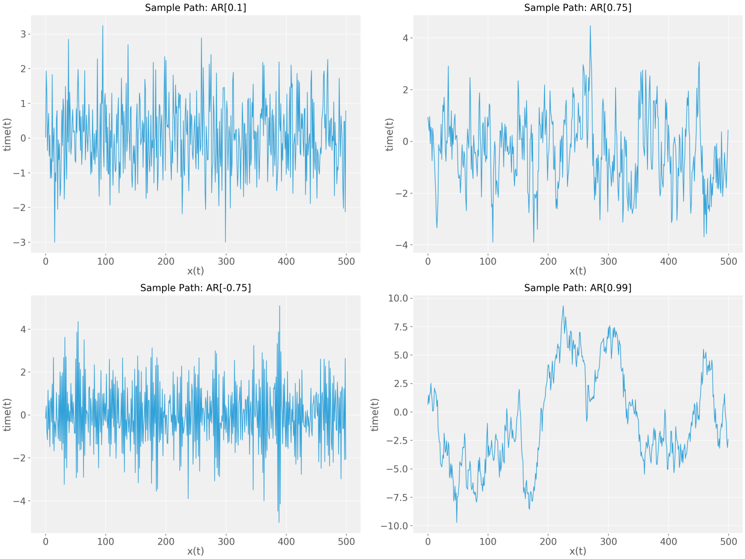
AR(1) Process

- Definition: **Auto****R**egressive Process of order 1, or **AR(1)**:

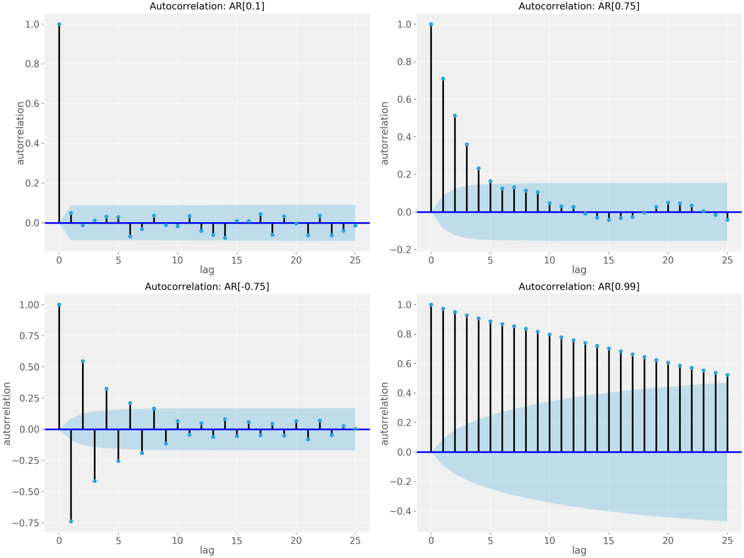
$$x_t = \mu + \phi x_{t-1} + \epsilon_t, \text{ where } \epsilon_t \sim \text{WN}(0, \sigma_\epsilon)$$

- Properties of an AR(1) process:
  - Mean $E[x_t]$
  - Variance $\text{Var}(x_t)$:
  - Autocovariance γ_j:
  - Autocorrelation ρ_j:
  - Partial autocorrelation α_j:
- Question: What happens if $|\phi| \geq 1$?

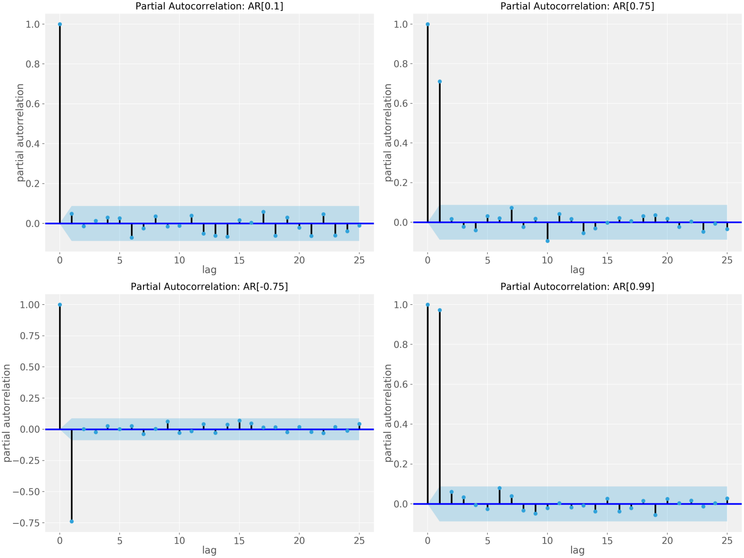
AR(1) WITH DIFFERENT ϕ: SIMULATIONS



AR(1) WITH DIFFERENT ϕ: SAMPLE AUTOCORRELATION FUNCTIONS



AR(1) with different ϕ: Sample partial autocorrelation functions



AR(1) FORECASTS

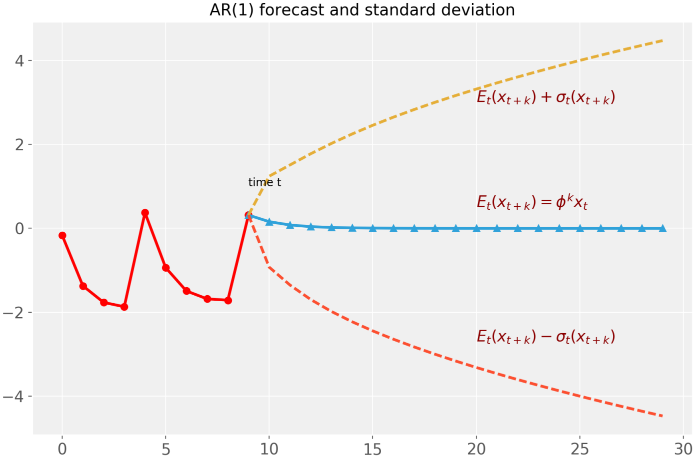
$$x_t = \phi x_{t-1} + \epsilon_t$$

Forecasts (for $\mu = 0$):

$$E_t[x_{t+k}] =$$

$$\text{Var}_t(x_{t+k}) =$$

AR(1) FORECASTS



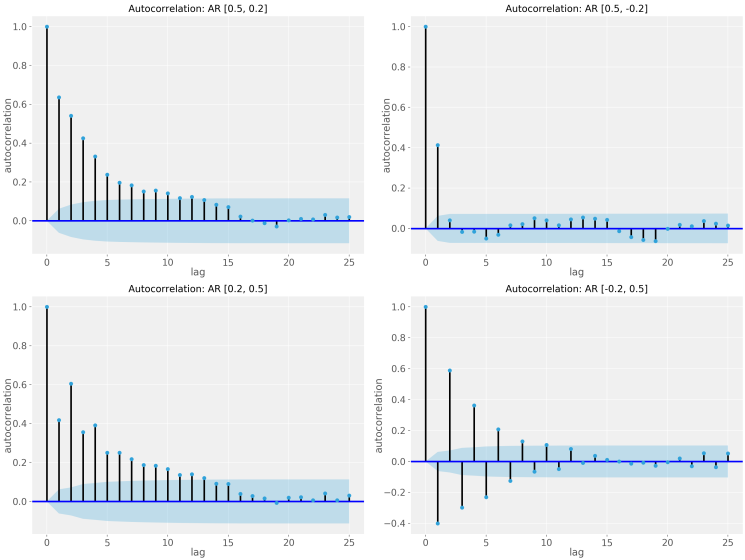
AR(p) PROCESSES

- Definition: AutoRegressive process of order p (AR(p)):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \epsilon_t$$

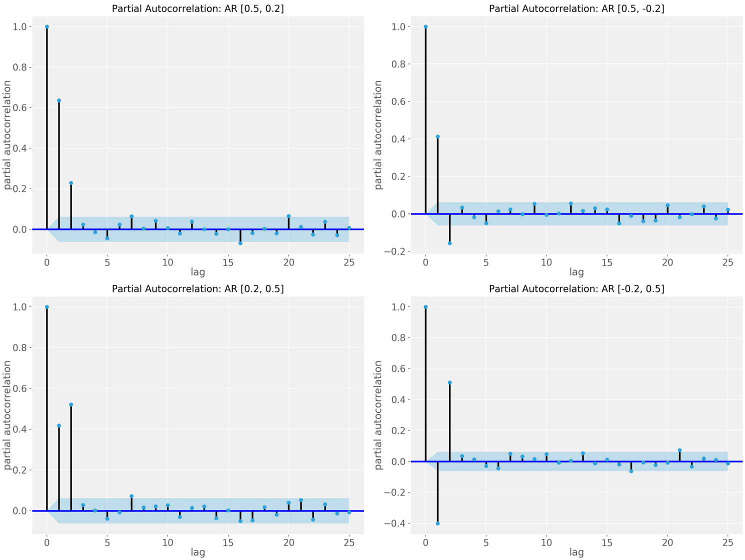
- Expressions for variances and autocovariances are more complicated than for AR(1)

AR(p): ACF



Autocorrelation: AR [0.5, 0.2]
Autocorrelation: AR [0.5, -0.2]
Autocorrelation: AR [0.2, 0.5]
Autocorrelation: AR [-0.2, 0.5]
autocorrelation
lag

AR(p): PACF


Partial Autocorrelation: AR [0.5, 0.2]
Partial Autocorrelation: AR [0.5, -0.2]
Partial Autocorrelation: AR [0.2, 0.5]
Partial Autocorrelation: AR [-0.2, 0.5]
partial autocorrelation
lag

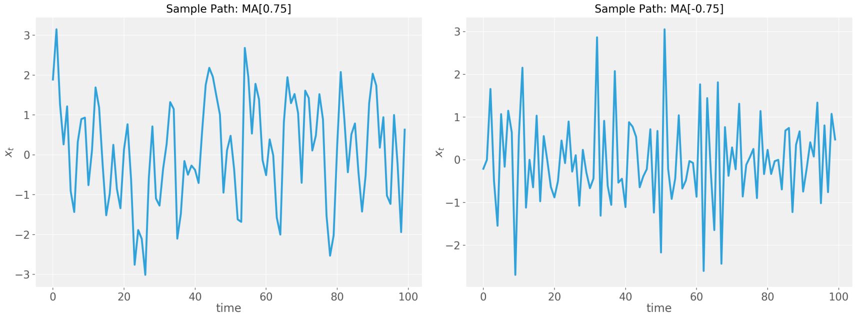
MOVING AVERAGE PROCESSES MA(q)

- Definition: Moving Average process of order q (MA(q)):

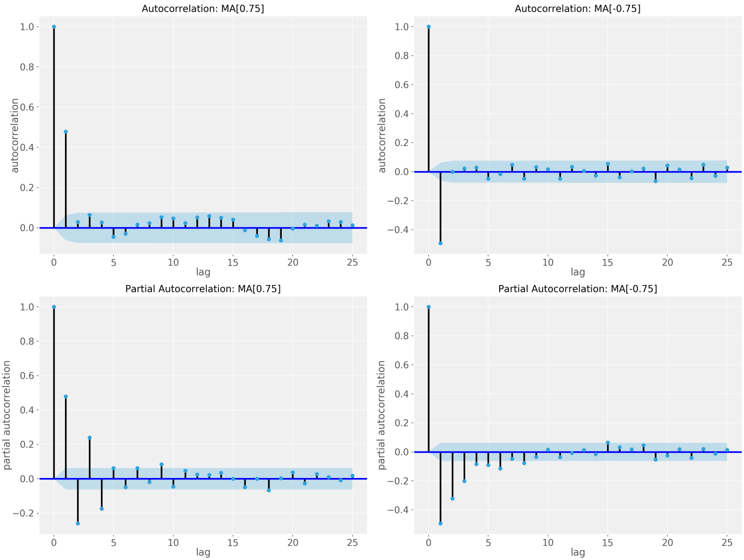
$$x_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

- Examples:
  - $\theta_j = 1/q$: lagging mean over last q periods
  - $\theta_j = z^j$: average geometrically declining weights
  - q can be infinity
- Properties of an MA(1) process:
  - Mean $E[x_t]$
  - Variance $\mathrm{Var}(x_t)$
  - Autocovariance γ_j

MA(1) WITH DIFFERENT θ: SAMPLE PATHS



MA(1) WITH DIFFERENT θ: ACF AND PACF FUNCTIONS

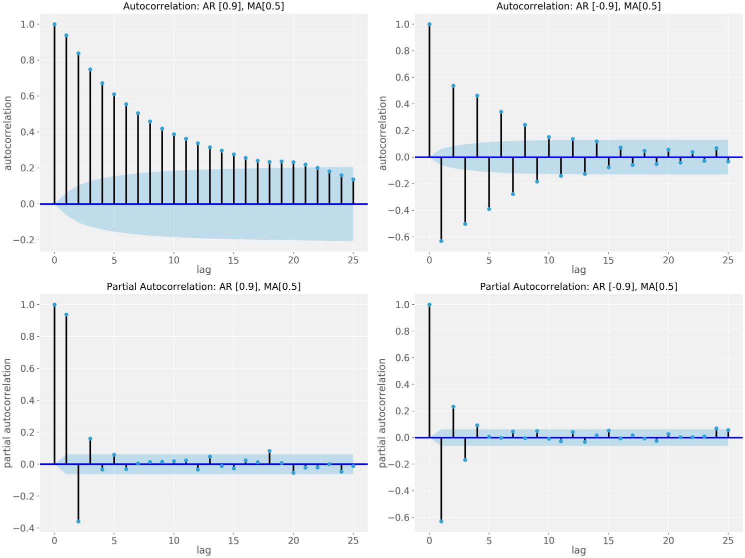


General ARMA process

General ARMA(p, q) process:

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}$$

ARMA(1,1) WITH DIFFERENT θ: ACF AND PACF FUNCTIONS



Some of you (e.g. engineers) might be familiar with spectral representations and Fourier transforms. ARMA models can easily be represented in the spectral domain. Chapter 8 in Cochrane's notes has a nice treatment. If you are used to thinking about spectral representations, this chapter might be useful.

If you are NOT familiar with spectral representations, you can ignore this altogether as we will not use these concepts in class.

USEFUL TOOL: LAG OPERATORS

- ARMA models can be manipulated using **lag operators**:

$$
\begin{aligned}
Lx_t &= x_{t-1} \\
L^2x_t &= LLx_t = Lx_{t-1} = x_{t-2} \\
L^jx_t &= x_{t-j} \\
L^{-j}x_t &= x_{t+j}
\end{aligned}
$$

- Definition: **Lag polynomial**

$$\Phi(L) = 1 + \phi L + \phi^2 L^2 + \ldots$$

- Definition: **Difference operator**

$$\Delta x_t \equiv x_t - x_{t-1} = (1-L)x_t$$

$$\Delta^n x_t \equiv (1-L)^n x_t$$

Lag operators: Examples

AR(1):

$$x_t = \phi x_{t-1} + \epsilon_t$$

$$\Rightarrow x_t - \phi x_{t-1} = \epsilon_t$$

$$\Rightarrow (1 - \phi L) x_t = \epsilon_t$$

$$\Rightarrow \Phi(L) x_t = \epsilon_t \quad \text{where} \quad \Phi(L) = 1 - \phi L$$

Generic AR:

$$\Phi(L) x_t = \epsilon_t$$

MA(1)

$$x_t = \epsilon_t + \theta \epsilon_{t-1}$$

$$\Rightarrow x_t = (1 + \theta L) \epsilon_t$$

$$\Rightarrow x_t = \Theta(L) \epsilon_t \quad \text{where} \quad \Theta(L) = 1 - \theta L$$

Generic MA:

$$x_t = \Theta \epsilon_t$$

- Generic ARMA:

$$\Phi(L)\,x_t = \Theta(L)\,\epsilon_t,$$

$$\text{where}\;\; \Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \ldots - \phi_p L^p$$

$$\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3 + \ldots + \theta_q L^q$$

Inverting an AR(1) to a MA(∞)

- By repeated substitution:

$$
\begin{aligned}
x_t &= \phi x_{t-1} + \epsilon_t \\
&= \phi^2 x_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \\
&= \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots
\end{aligned}
$$

- AR(1) can be represented as MA(∞)
- General method using lag operators

$$
\begin{aligned}
(1 - \phi L)\, x_t &= \epsilon_t \\
x_t &= (1 - \phi L)^{-1} \epsilon_t \\
&= \left(1 + \phi L + \phi^2 L^2 + \dots\right) \epsilon_t \\
&= \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}
\end{aligned}
$$

- Valid only if $\left|\phi\right| < 1$!

- Inversion of higher order AR(p) : Cochrane ch. 3.3.4
- Factorization of an AR(2):

$$\left(1-\phi_1 L-\phi_2 L^2\right) x_t=\epsilon_t$$

$$\left(1-\lambda_1 L\right)\left(1-\lambda_2 L\right) x_t=\epsilon_t,$$

$$\text{where } \lambda_1 \lambda_2=-\phi_2$$

$$\lambda_1+\lambda_2=\phi_1.$$

- Any AR(p) process can be factorized with $\lambda_1, \ldots, \lambda_p$.
- The **roots** of an AR(p) are $1/\lambda_i$.

General result: Sums of ARMA are ARMA!

Sums of ARMA(p, q) (Hamilton ch. 4.7):

- MA(1) + white noise = MA(1)
- MA(q) + MA(p) = MA(max(p, q))
- AR(p_1) + AR(p_2) = ARMA($p_1 + p_2$, max(p_1, p_2))
- ARMA(p_1, q_1) + ARMA(p_2, q_2) = ARMA($p_1 + p_2$, max($p_1 + q_2, p_2 + q_1$))

Invert an MA(1):

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1}$$

What are the AR representations for Δx_t and $\Delta^2 x_t$?

$$x_t = 2x_{t-1} - x_{t-2} + \epsilon_t$$

Derive the MA representation of an AR(2):

$$(1 - L)(1 - L)x_t = \epsilon_t$$

- The behavior of ARMA processes can be analyzed by using their ACF and PACF functions
- Another useful tool is the **Impulse Response Function (IRF)**
- Idea: Suppose a single shock to ϵ_t occurs in period t, all other $\epsilon_s = 0, s \neq t$
- The IRF tells us how this single shocks affects $x_{t+1}, x_{t+2}, x_{t+3}, \ldots$

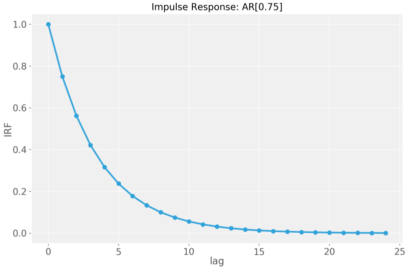
IRFs

Recall: The MA representation of an ARMA process is

$$x_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \ldots$$

θ_j measures the impact of a shock j periods ago on x_t.

$\Rightarrow$ The **Impulse Response Function** traces the effect of a shock in t on x_{t+j}:



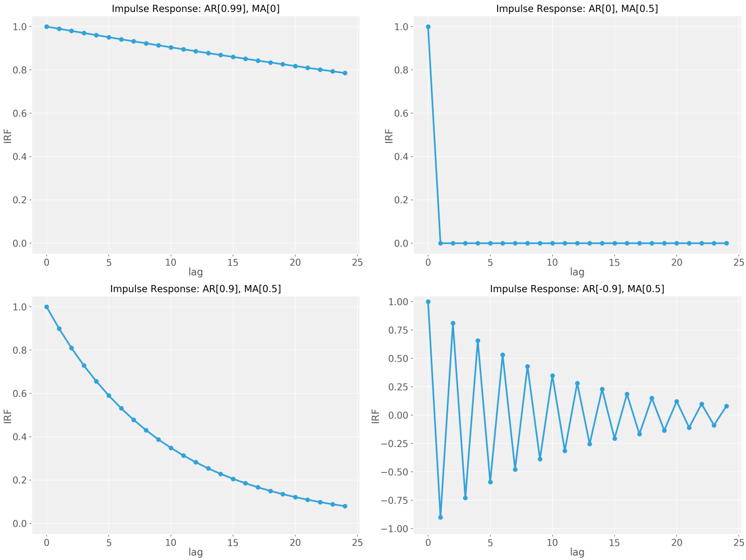
Half-life of a shock: Lag j such that $\text{IRF}_j = 0.5$

Half-life of an AR(1)

$$\phi^j = 0.5 \Rightarrow j = -\frac{\log 2}{\log \phi}$$

ϕ	0.250	0.500	0.750	0.900	0.950	0.990	0.999
HL:	0.500	1.000	2.409	6.579	13.513	68.968	692.801

Impulse response functions of ARMA processes



Note: All IRFs in these examples converge to 0!

Summary: ARMA model

- An ARMA(p, q) process is defined

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}$$

- Lag operators:

$$\Phi(L)\, x_t = \Theta(L)\, \epsilon_t,$$

$$\text{where} \quad \Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \ldots - \phi_p L^p$$

$$\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3 + \ldots + \theta_q L^q$$

- AR and MA processes can be inverted
- ACFs and PACFs depend on AR and MA parameters
- IRFs measure impact of a shock on future values of the time series
- Our examples: Conditional forecasts and variances converge to unconditional moments, IRFs converge to zero

Outline

Stationarity and ergodicity

Somewhat technical but VERY important!

Recall AR(1) process: $x_t = \phi x_{t-1} + \epsilon_t$:

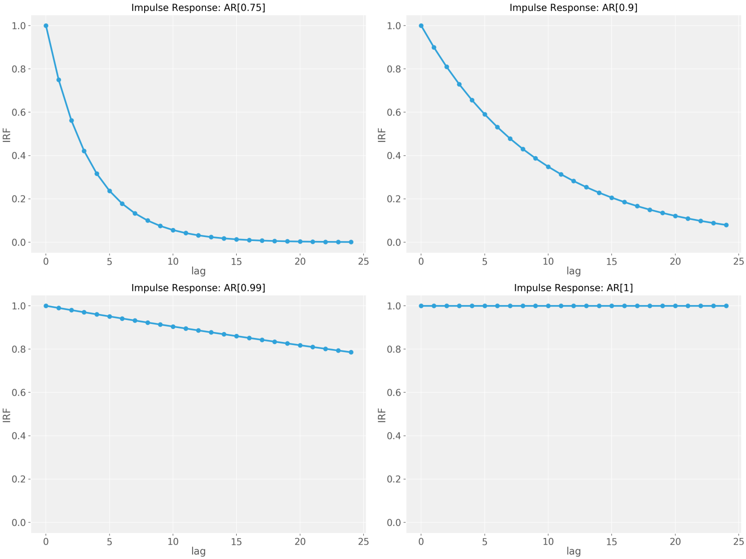
If $|\phi| < 1$:

$$\lim_{j\to\infty} E_t[x_{t+j}] = 0 = E[x_t]$$

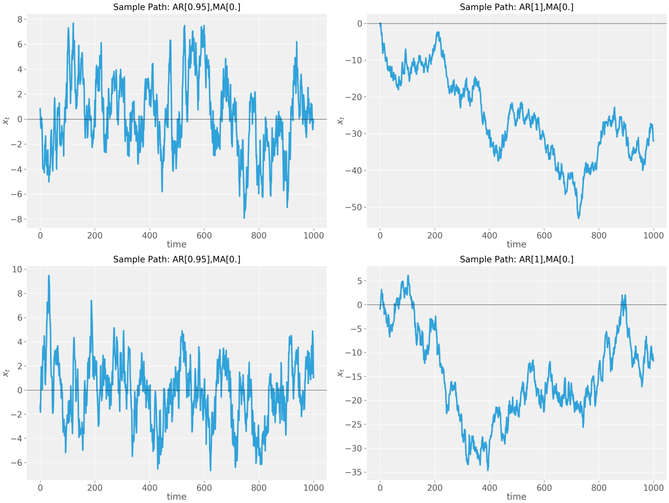
$$\lim_{j\to\infty} \text{Var}_t(x_{t+j}) = \frac{\sigma_\epsilon^2}{1-\phi^2} = \text{Var}(x_t)$$

What happens if $|\phi| \geq 1$?

IRFs


Impulse Response: AR[0.75]
Impulse Response: AR[0.9]
Impulse Response: AR[0.99]
Impulse Response: AR[1]
IRF
lag
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25

2 SAMPLE PATHS FOR $\phi = 0.95$ AND $\phi = 1$



Note the y-scales!

Intuition:

- **Stationarity**: The distribution of the stochastic process remains unchanged over time.
- **Ergodicity**: The past does not affect the future "too strongly".

Time series that are stationary and ergodic are easy to estimate and have desirable properties.

Time series that are non-stationary/non-ergodic require a different set of techniques.

Important lesson: Make sure your time series is stationary and ergodic!!

Note: Often "stationary" means "stationary and ergodic".

Definition 4 (Strict (or strong) stationarity).

A process is **strictly stationary** if, for any values of $j_1, j_2, \ldots j_n,$ the joint distribution of $\left(x_t, x_{t+j_1}, ..x_{t+j_n}\right)$ depends only on the intervals separating the dates $(j_1, j_2, \ldots j_n)$ and not on the dates themselves (t).

- Strict stationarity requires that the joint distribution of a stochastic process does not depend on time
- The only factor affecting the relationship between two observations is the gap between them.
- Strict stationarity is weaker than i.i.d. since the process maybe serially correlated but it is nonetheless a strong assumption and sometimes violated in financial and macroeconomic data.

Definition 5 (Covariance (or weak) stationarity).

A stochastic process x_t is **covariance (or weakly) stationary** if

$$
\begin{aligned}
E(x_t) &= \mu < \infty && \forall t \\
\mathrm{Var}(x_t) &= \sigma^2 < \infty && \forall t \\
\mathrm{Cov}(x_t, x_s) &= \gamma(|t-s|) && \forall t, s
\end{aligned}
$$

- Covariance stationarity requires that both the unconditional mean and unconditional variance are finite and do not change with time.
- Note that covariance stationarity only applies to unconditional moments and not conditional moments, and so a covariance process may have a varying conditional mean (i.e. be predictable).

- Strict stationarity plus finite first and second moment $\Rightarrow$ weak stationarity
- Covariance stationarity does NOT imply strict stationarity
- Covariance stationarity plus normality $\Longleftrightarrow$ strict stationarity

- Strong stationarity is useful for proving some theorems.
- For example, a nonlinear function of a strongly stationary variable is strongly stationary; this is not true of covariance stationarity.
- For most purposes, weak or covariance stationarity is sufficient.
- **Unless otherwise noted, "stationarity" refers to "covariance stationarity".**

Ergodicity

- In a strictly stationary or covariance stationary stochastic process, no assumption is made about the strength of dependence between random variables in the sequence.
- For example, in a covariance stationary stochastic process it is possible that

$$\rho_1 = \text{Corr}(x_t, x_{t-1}) = \rho_{100} = \text{Corr}(x_t, x_{t-100}) = 0.5$$

- **Ergodicity**: The strength of dependence between random variables in a stochastic process diminishes the farther apart they become and eventually $\rho_j \to 0$

Definition 6 (Ergodicity (intuitive definition)).

A stochastic process $\{x_t\}$ is **ergodic** if any two collections of random variables partitioned far apart in the sequence are "essentially" independent.

Intuition: The stochastic process $\{x_t\}$ is ergodic if x_t and x_{t-j} are "essentially uncorrelated" if j is large enough.

Theorem 1.

If the autocovariance function of a covariance-stationary process satisfies

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty$$

then the process is ergodic.

Question: What is the LLN for i.i.d. random variables?

Next: LLN for TS that are serially correlated but stationary and ergodic but not i.i.d.?

The **ergodic theorem** states that the LLN holds even when a stochastic process is serially correlated:

Theorem 2 (Ergodic theorem).

If $\{x_t\}$ *is* ***stationary and ergodic****, then*

$$\bar{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t \longrightarrow E[x_t] \ \ as \ \ T \longrightarrow \infty.$$

Implication: The **Method of Moments** is applicable **as long as the stochastic process is stationary and ergodic**.

Unless you have a degree in Statistics, most of what you have learned requires stationarity and ergodicity!

Recall AR(1) process:

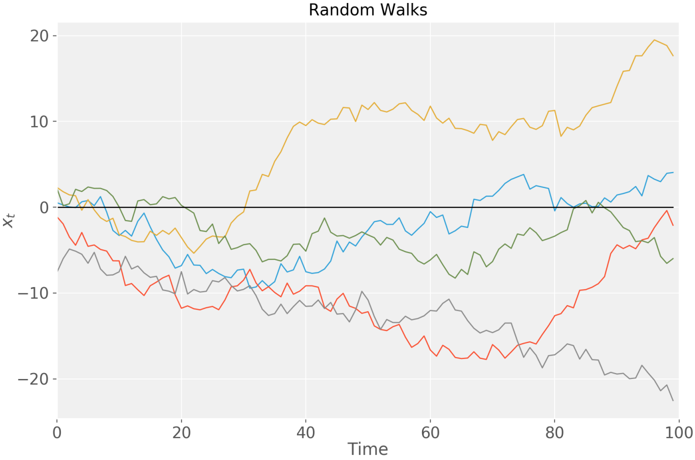
$$\begin{aligned} x_t &= \phi x_{t-1} + \epsilon_t \\ &= \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \dots \\ E[x_t] &= 0 \ \ \forall t \\ \text{Var}(x_t) &= \frac{\sigma_\epsilon^2}{1-\phi^2} \ \ \forall t \\ \gamma_j &= \phi^j \ \ \forall t \end{aligned}$$

Clearly, x_t is stationary and ergodic if $|\phi| < 1$.

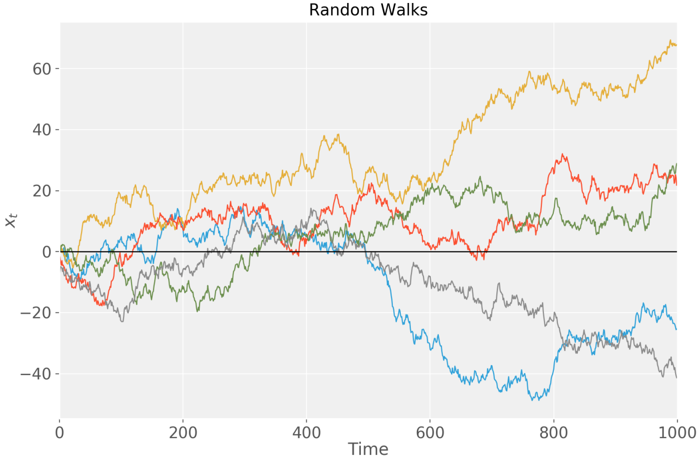
A random walk is an AR(1) with $\phi = 1$:

$$x_t = x_{t-1} + \epsilon_t$$
$$= \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \ldots + x_0$$

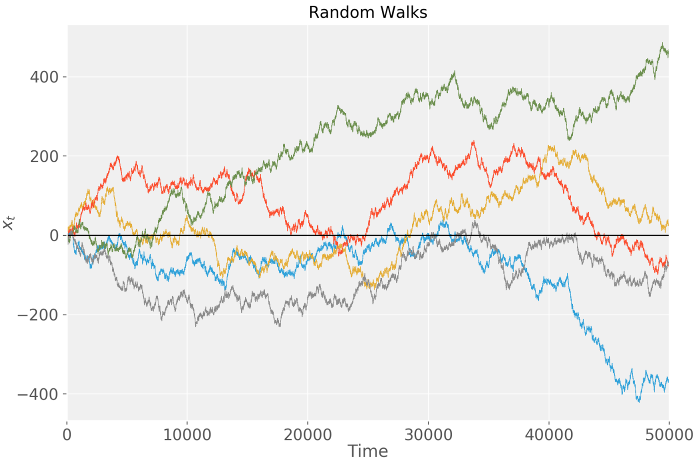
Question: Is a random walk weakly/strictly stationary and/or ergodic?


Random Walks
20
10
0
−10
−20
x_t
0
20
40
60
80
100
Time

Simulations of random walk: T=1,000

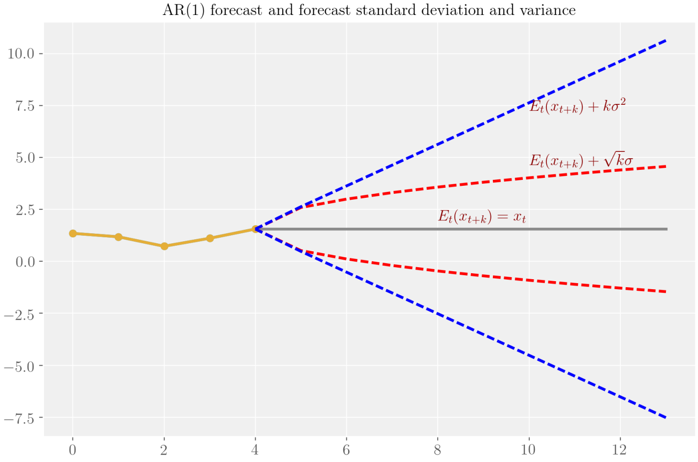


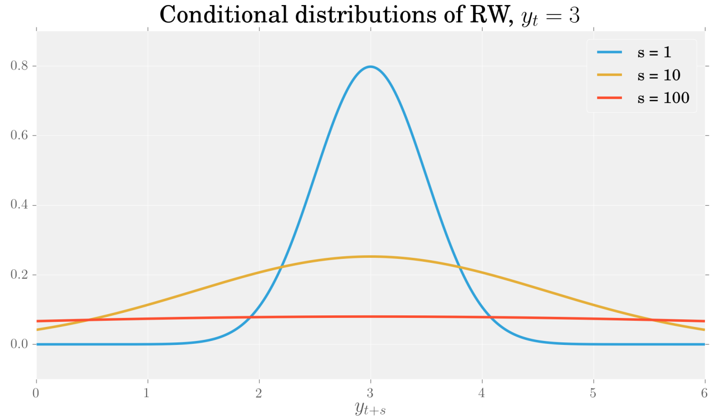
Note the *y*-scale!



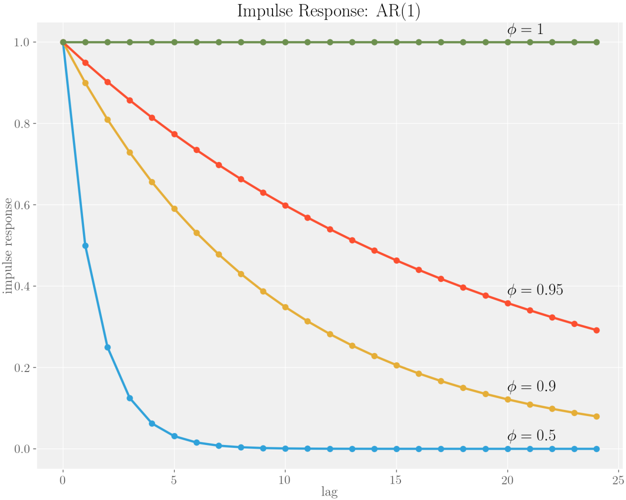
Note the *y*-scale!

Random walk forecasts




Conditional distributions of RW, $y_t = 3$
s = 1
s = 10
s = 100
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
y_{t+s}

Impulse responses



Note: IRF does not converge to 0 for $\phi = 1 \rightarrow$ A RW is not ergodic

- AR(1) :

$$x_t = \phi x_{t-1} + \epsilon_t \;\; \Rightarrow \;\; |\phi| < 1$$

- AR(p) :

$$\left(1 - \phi_1 L - \ldots - \phi_p L^p\right) x_t = \epsilon_t$$

$$\Leftrightarrow \left(1 - \lambda_1 L\right) \ldots \left(1 - \lambda_p L\right) x_t = \epsilon_t$$

$$\Rightarrow |\lambda_i| < 1 \;\; \forall i$$

$\longrightarrow$ All roots of polynomial are outside the unit circle.

- MA(q) :

$$x_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

$$E[x_t] = \mu$$

$$\text{Var}(x_t) = \sigma_\epsilon^2 \left(1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2\right)$$

Any (finite-order) MA(q) is covariance stationary (and ergodic) if $\sum_{j=0}^{\infty} \theta_j^2 < \infty$
$\Rightarrow$ any (finite-order) MA(q) is stationary

Definition 7.

For any finite order ARMA(p, q)

$$(1-\lambda_1 L)\ldots(1-\lambda_p L)x_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q}:$$

- If n of the $\lambda_i = 1$, we say that x_t is **integrated** of order n: $I(n)$
- If one $\lambda_i = 1$, x_t has a **unit root**
- If $x_t \sim I(1)$ but $\Delta x_t \sim I(0)$, then x_t is **difference stationary**
- An ARMA(p, q) process with n unit roots is called ARIMA(p, n, q)

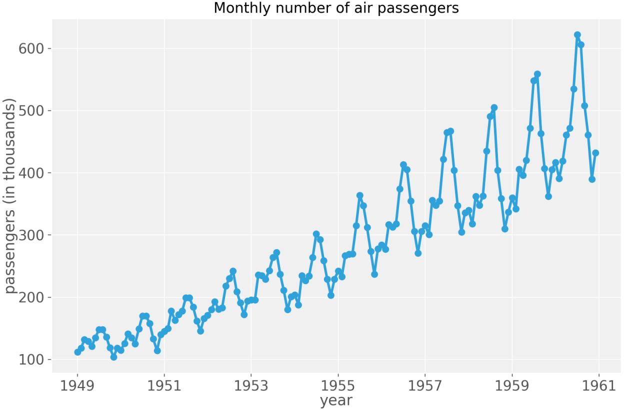
Is x_t stationary?

$$x_t = \frac{3}{2}x_{t-1} - \frac{1}{2}x_{t-2} + \epsilon_t$$

Are x_t and/or Δx_t stationary?

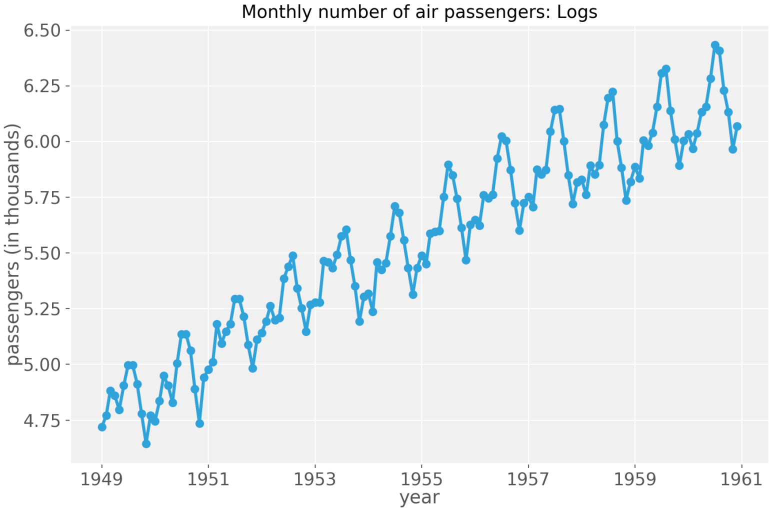
$$x_t = 2x_{t-1} - x_{t-2} - \epsilon_t$$

Dealing with non-stationary time series: Air passengers

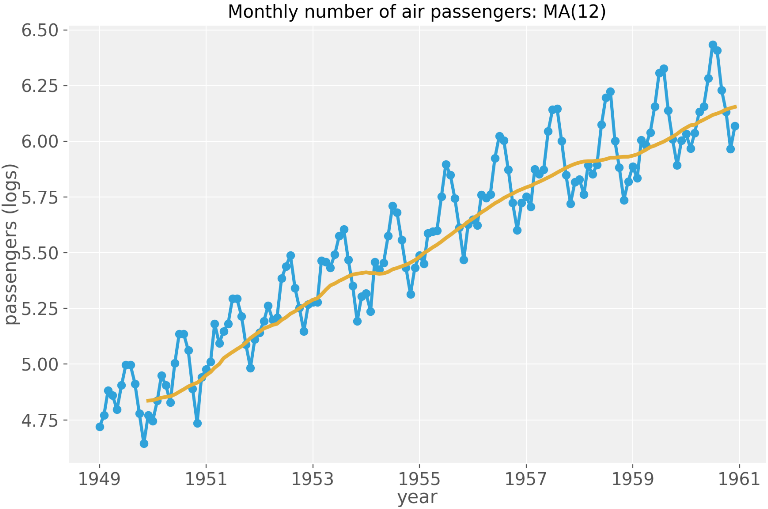


Is this time series stationary?

Example of nonstationary time series: Logs



EXAMPLE OF NONSTATIONARY TIME SERIES: MA(12)



A simple decomposition of a time series is:

$$x_t = T_t + S_t + e_t,$$

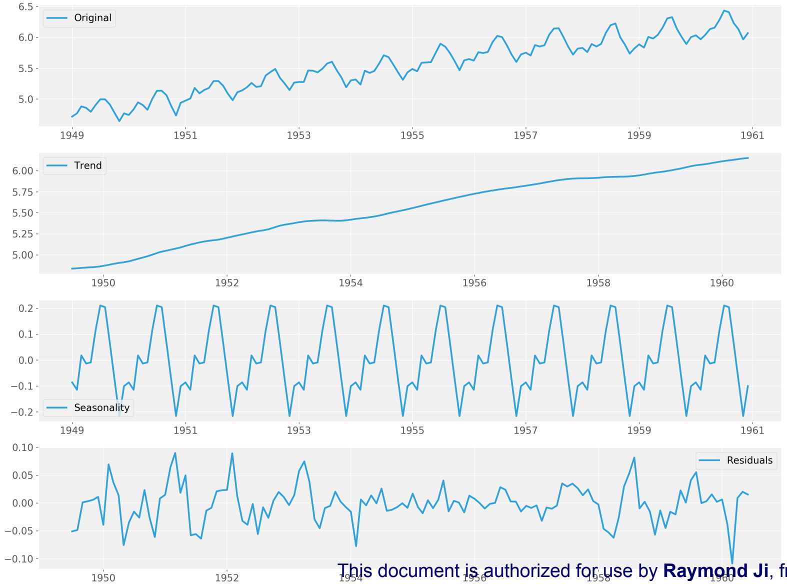
where

- T_t is a deterministic trend
- S_t is a seasonal component
- e_t is the residual that is (hopefully) stationary.

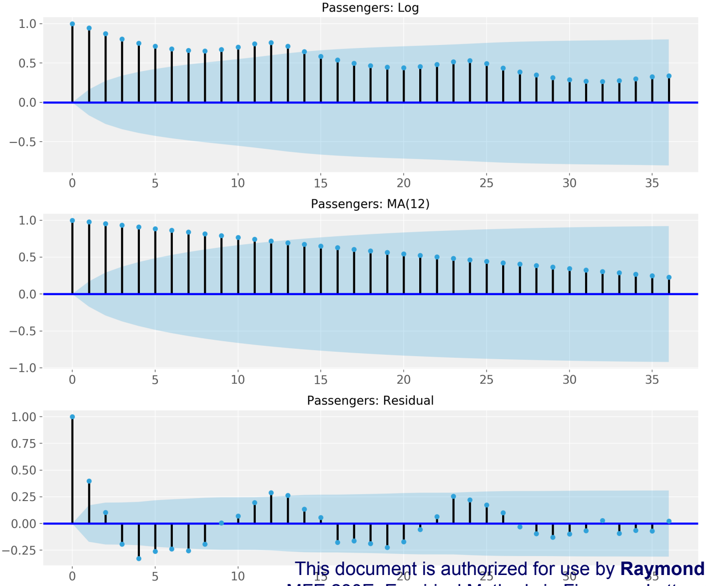
See Ruppert ch. 13.1-2 for details. In `statsmodels`:

```
from statsmodels.tsa.seasonal import seasonal_decompose
decomposition = seasonal_decompose(airpg_df.p_log)
trend = decomposition.trend
seasonal = decomposition.seasonal
residual = decomposition.resid
```

Example of nonstationary time series: Seasonal decomposition



Example of nonstationary time series: ACF



Dealing with $I(1)$ processes in practice

Suppose your times series is a random walk:

$$x_t = x_{t-1} + \epsilon_t$$

Dealing with x_t directly has many econometric challenges.

Solution: Take the first difference of Δx_t since $\Delta x_t = \epsilon_t$ is stationary.

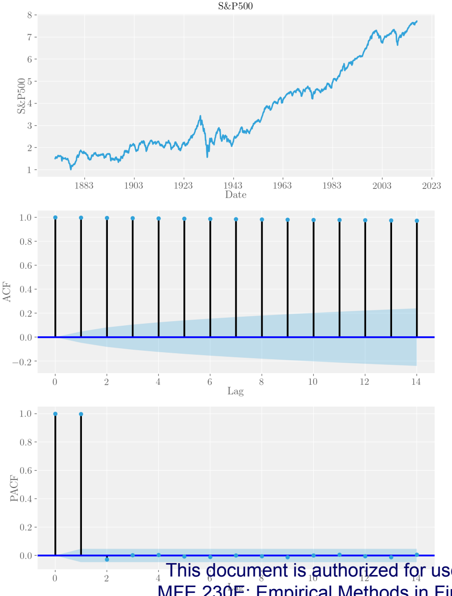
Note: Taking Δ^2 of x_t leads to MA terms ("over-differencing")

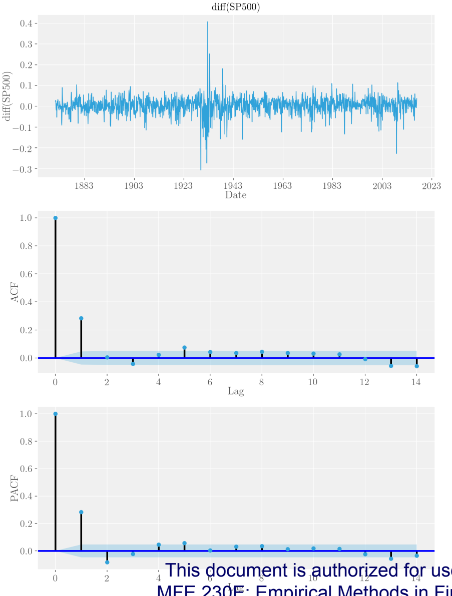
$$\Delta^2 x_t = (1 - L)\epsilon_t = \epsilon_t - \epsilon_{t-1}$$

DEALING WITH $I(n)$ PROCESSES IN PRACTICE

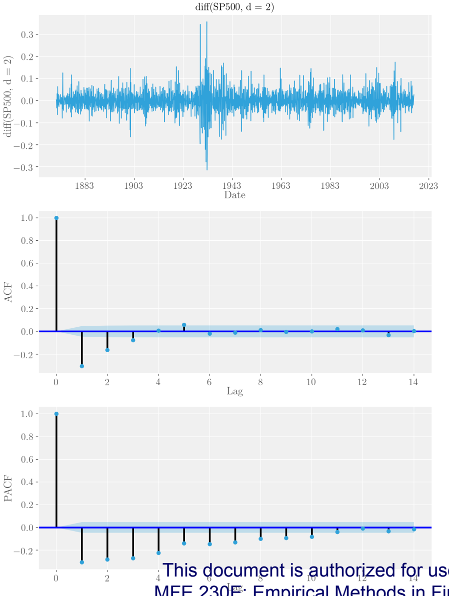
1. Determine order of ARIMA(p, n, q) (Next week)
2. Take n differences: $\Delta^n x_t \sim \text{ARMA}(p, q)$
3. Make sure not to "under-difference": $\Delta^m x_t$ where $m < n$. Why?
4. Make sure not to "over-difference": $\Delta^m x_t$ where $m > n$. Why?

Example: Level of (log) S&P 500 index

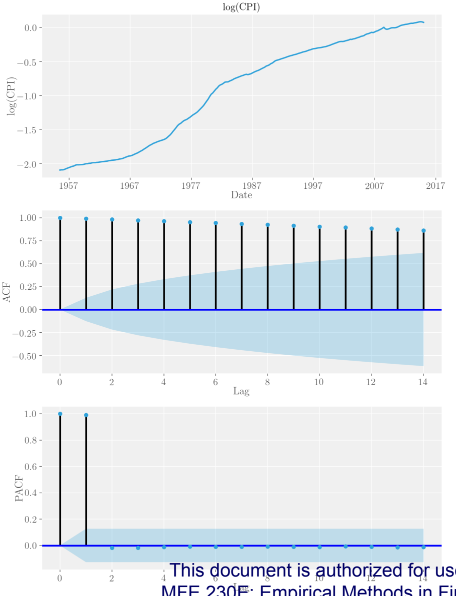



diff(SP500)
diff(SP500)
Date
1883
1903
1923
1943
1963
1983
2003
2023
ACF
Lag
PACF
Lag

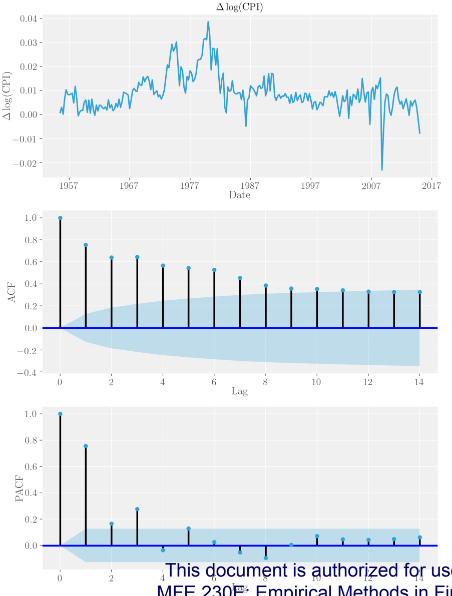
Example: Second differences of (log) S&P 500 index



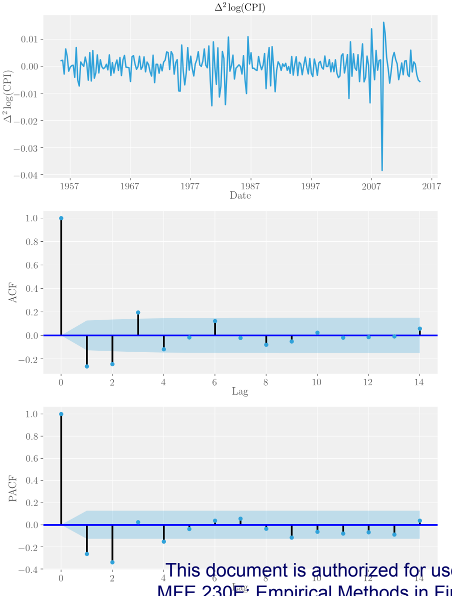
Example: Level of (log) CPI (consumer price index)



Example: First differences of CPI



EXAMPLE: SECOND DIFFERENCES OF CPI



S&P 500: Clearly working with (log) first differences is the best choice

CPI: Choice between first and second differences of (log) CPI is less clear

⇒ Need formal statistical methods

Wold decomposition theorem

- Stationary and ergodic ARMA processes are an important class of linear time series models.
- Question: How "general" is this class?
- It turns out that every covariance stationary process (even nonlinear ones) has a linear ARMA representation!

Theorem 3 (Wold Decomposition Theorem).

Any purely stochastic ***covariance stationary process*** $\{x_t\}$ *can be represented as*

$$x_t = \sum_{j=0}^{\infty} b_j u_{t-j}$$

where u_t *is the* ***linear*** *forecast error of* x_t *given information in lags of* y_t. *Furthermore,* $b_0 = 1$, $\sum_{j=0}^{\infty} b_j^2 < \infty$ *and the* $\{b_j\}$ *and* $\{u_s\}$ *are unique.*

Implication: If two processes have the same Wold representation, they are identical.

Note: The Wold shocks are not the 'true' shocks in a nonlinear model.

Let $g(.)$ be a nonlinear function. Then

$$\begin{aligned} x_t &= g(e_t, e_{t-1}, e_{t-2}, \dots) \\ &= \sum_{j=0}^{\infty} b_j u_{t-j} \\ \Rightarrow e_t &\neq u_t \end{aligned}$$

- **Stationarity**: The distribution of the stochastic process remains unchanged.
- **Ergodicity**: The past does not affect the future "too strongly".
- Covariance stationarity: $\text{Corr}(x_t, x_s) = \gamma(|t - s|) \quad \forall t, s$
- If $\sum_{j=0}^{\infty} |\gamma_j| < \infty$, then the process is ergodic
- An AR(1) $x_t = \phi x_{t-1} + \epsilon_t$ is stationary and ergodic iff $|\phi| < 1$
- If $\phi = 1$, the process has a unit root
- An AR(p) is stationary and ergodic if all roots $|\lambda_i| < 1$
- Any finite MA is stationary and ergodic

Next week

- Estimation of ARMA processes
- Why the "classical" linear regression model is not valid for time series

  Before class: Review lectures 8 and 9 of pre-program Stats class!