# Lecture 1: Characteristics and Examples of Time Series Data

Introduction to Time Series, Fall 2024

Ryan Tibshirani

Related reading: Chapters 1.1–1.2 of Shumway and Stoffer (SS); Chapters 2.3 and 3.2–3.6 of Hyndman and Athanasopoulos (HA).

## 1 Course stuff

- Instructor: Ryan Tibshirani
- GSI: Tiffany Ding
- Reader: Theo Pan
- Course website: https://stat153.berkeley.edu/fall-2024/
- Everything will be up on the website: lecture notes, homework assignments, syllabus, schedule, links to bCourses, Ed discussion, etc.
- Please email the GSI with any issues first. The Instructor will be looped in only as-needed
- Please call me Ryan, Professor Tibshirani, or Professor Tibs. Please do not call me Professor
- There will be 5 homework assignments, 1 midterm, and 1 final exam. Syllabus gives details on grading breakdown
- The homeworks will be due about every 2 weeks, with spacing for the midterm and the final. Schedule on website gives projected dates
- Probability at the level of Stat 134 or Data 140 is required as a pre-req. Statistics at the level of Stat 133 and 135 is recommend and may be taken concurrently
- We will also assume basic level of fluency in R programming. You will need to have R installed, and it will be very helpful for you to have RStudio installed
- Read the course website or the syllabus for the projected list of topics that we will cover
- Read the syllabus for late policy for homeworks, and collaboration policy
- Do not copy or cheat. It will not end well and dealing with it is really not fun for anyone involved
- Ok, now on to the fun stuff!

## 2 Time series intro

- What fields do time series data occur in? Economics, social science, epidemiology, medicine, neuroscience, language modeling, ...
  - Economics: stock prices or stock returns over time
  - Social science: birth rates or school acceptance rates over time
  - Epidemiology: Covid-19 cases or Influenza hospitalizations over time
  - Medicine: blood antibody levels over time (IgA, IgG, IgM, ...)
  - Neuroscience: brain-wave patterns over time, under different conditions

  - Language modeling: word or token distributions over time

- What distinguishes time series from traditional (batch) data problems?

  *The data are not i.i.d. (independent and identically distributed). There is correlation induced by the fact that we are making observations over time.*

- Ignoring these correlations is going to be problematic. Enter time series analysis, models, and forecasts

- Worth mentioning at the outset that there are two view in classical time series analysis: *time domain* and *frequency domain* (also called *spectral*) approaches.

  - Time domain: language/tools for studying lagged relationships—e.g., what happened yesterday will influence today and tomorrow

  - Frequency domain: language/tools for studying study seasonality and cycles

- These are not mutually exclusive. We will mostly focus on the former (time domain approaches), but will briefly introduce the latter (frequency domain approaches) a bit later in the course

- We will also use a significant chunk of the course to emphasize the predictive perspective: forecasting, practical considerations therein, and important related topics like calibration and ensembling

## 3 Time series examples

- We'll step through the following examples, in Figures 1–7, and discuss each, including why the data cannot really be i.i.d.

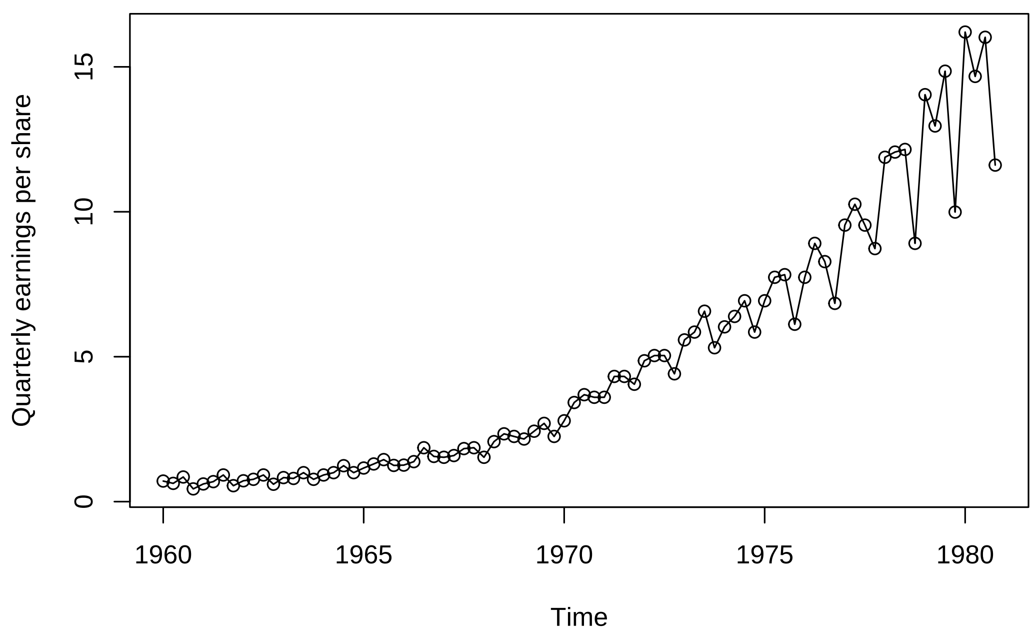


Figure 1: *Johnson & Johnson quarterly earnings per share (from SS).*

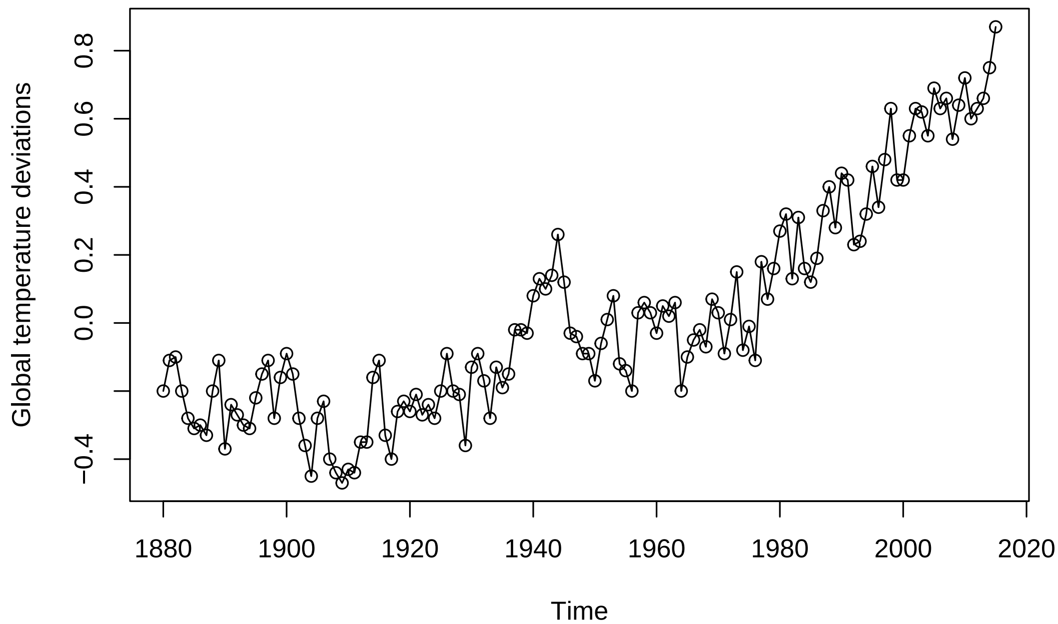


Figure 2: *Yearly average global temperature deviations from the 1951–1980 average (from SS).*

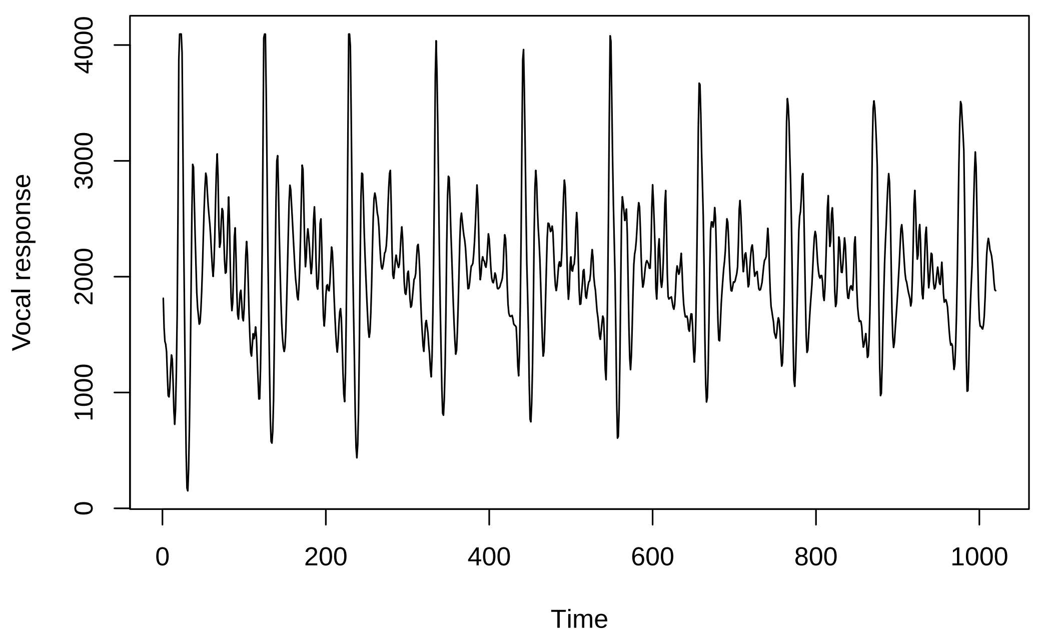


Figure 3: *Vocal response data measured from the syllable "aaa $\cdots$ hhh" (from SS).*

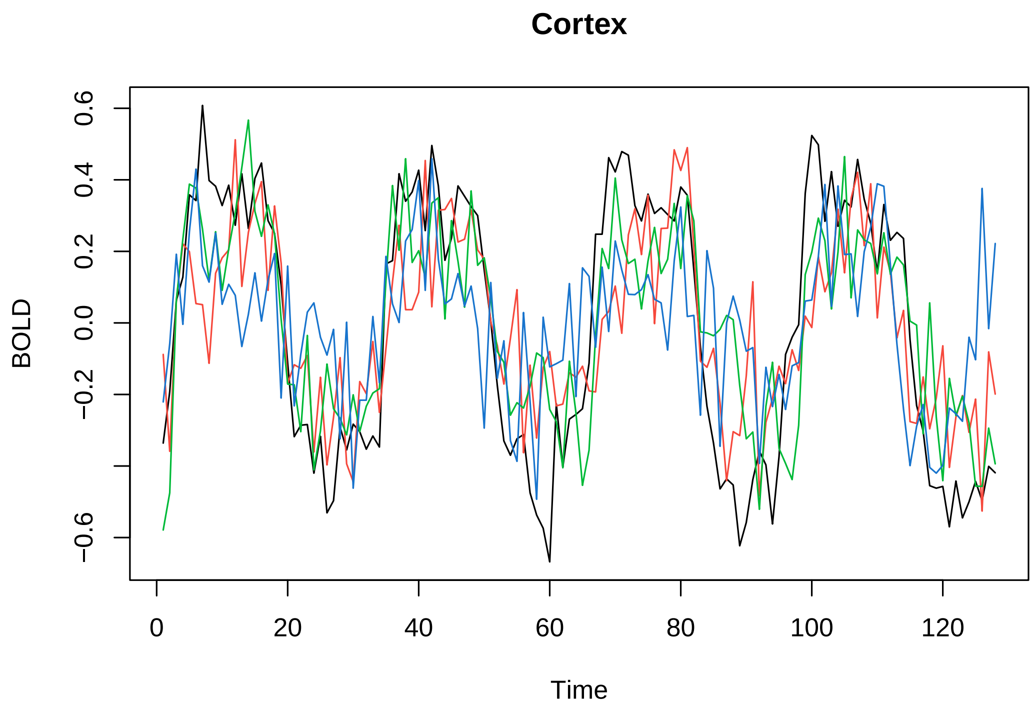


Figure 4: *Blood oxygenation-level dependent (BOLD) signal intensity in regions of the cortex (from SS).*

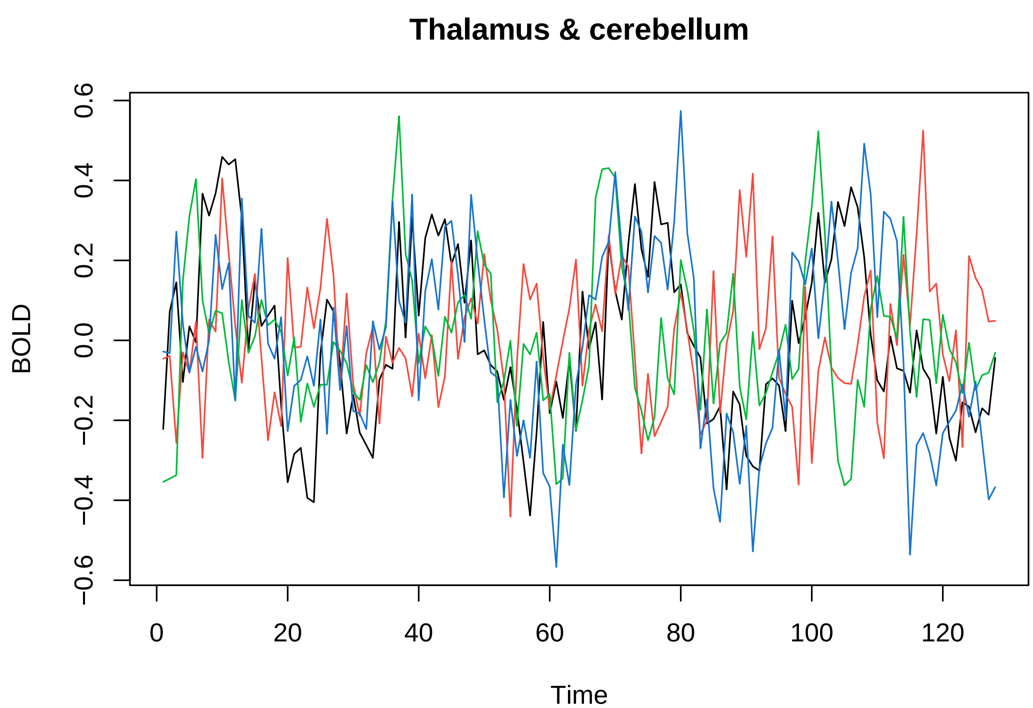


Figure 5: *BOLD signal intensity in regions of the thalamus and cerebellum (from SS).*

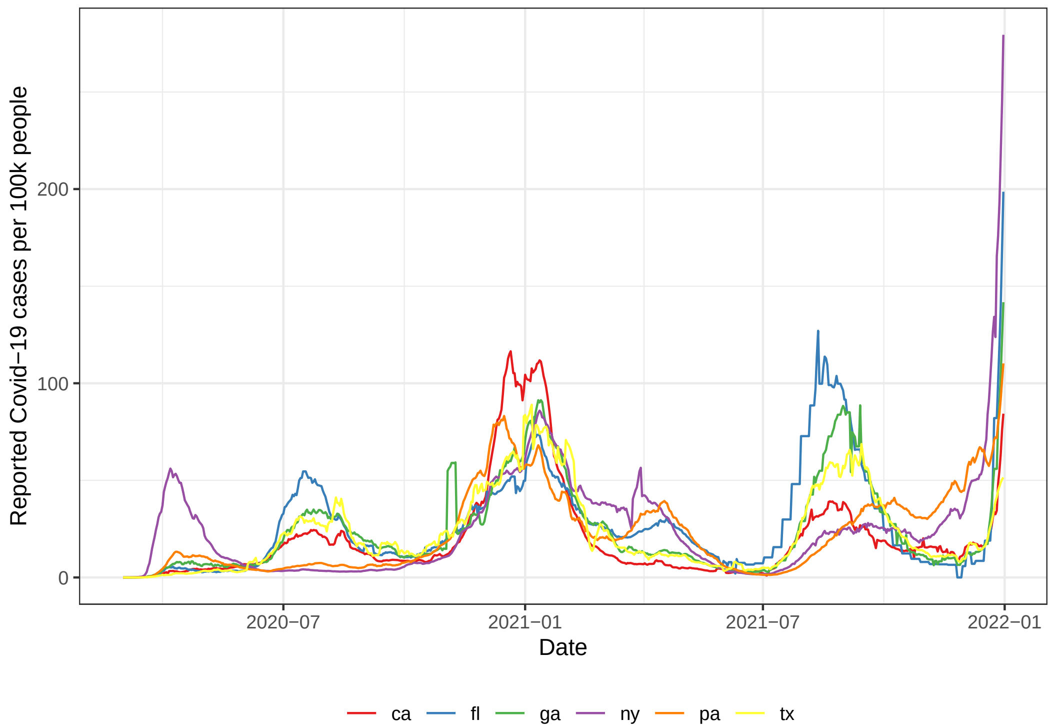


Figure 6: *Reported Covid-19 cases per 100k people in 6 large US states.*

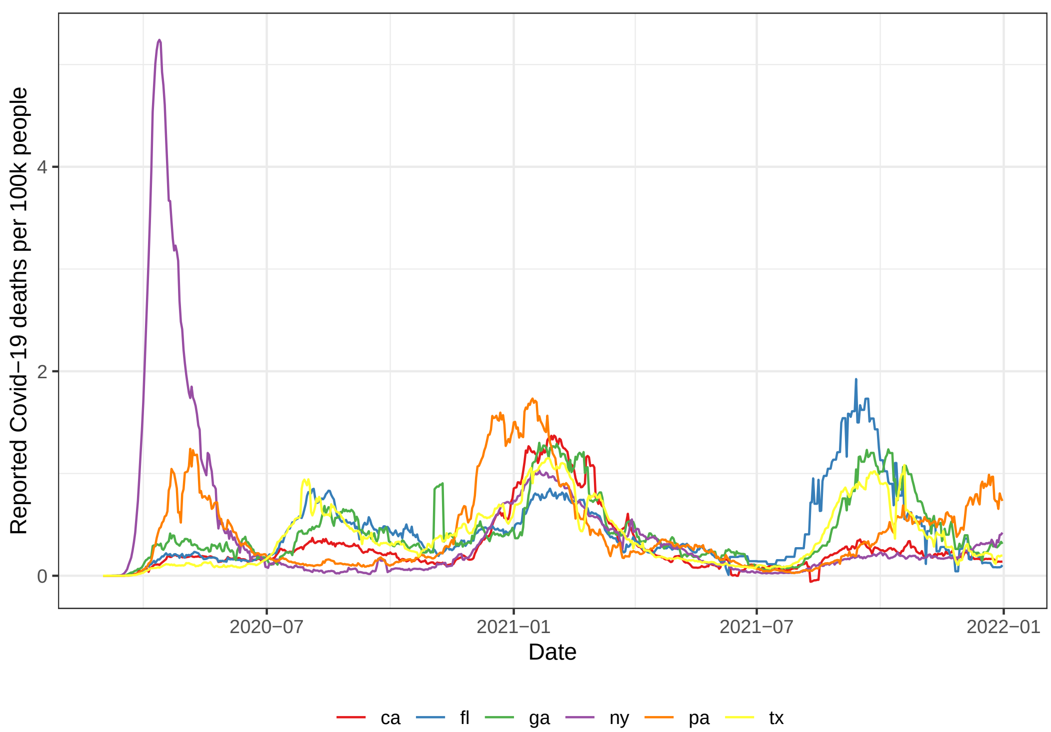


Figure 7: *Reported Covid-19 deaths per 100k people in 6 large US states.*

# 4 White noise

- The term *white noise* is used a lot in time series in related fields. It simply refers to a sequence $x_t$, $t = 1, 2, 3, \ldots$ of *uncorrelated* random variables, with zero mean, and constant variance. Precisely,

$$\mathrm{Cov}(x_s, x_t) = 0, \quad \text{for all } s \neq t$$
$$\mathbb{E}(x_t) = 0, \ \mathrm{Var}(x_t) = \sigma^2, \quad \text{for all } t$$

- (Recall ... for random variables $x, y$, their covariance is

$$\mathrm{Cov}(x, y) = \mathbb{E}\Big[(x - \mathbb{E}[x])(y - \mathbb{E}[y])\Big]$$

  and $\mathrm{Cov}(x, x) = \mathrm{Var}(x)$. The correlation between $x, y$ is

$$\mathrm{Cor}(x, y) = \frac{\mathrm{Cov}(x, y)}{\sqrt{\mathrm{Var}(x)}\sqrt{\mathrm{Var}(y)}}$$

  Therefore zero correlation and zero covariance are equivalent properties)
- A stronger property than white noise is *i.i.d. white noise*, that is, a sequence of white noise whose elements are also i.i.d.
- Why is this stronger? First, the distributions of the elements in a white noise sequence do *not* need to be the same—they only need to have the same first two moments (mean and variance). Second, white noise requires only zero correlation, not independence
- (Can you give an example of uncorrelated but not independent random variables?)
- An even stronger property is *Gaussian white noise*, that is, a sequence of white noise whose elements are also jointly Gaussian distributed
- Why is this stronger than i.i.d. white noise? Because if two Gaussians have equal mean and variance, then they are the same distribution; and, for Gaussians, zero correlation implies independence
- To summarize,

$$\{\text{Gaussian white noise sequences}\} \subseteq \{\text{i.i.d. white noise sequences}\} \subseteq \{\text{white noise sequences}\}$$

- A Gaussian white noise sequence is plotted below, in Figure 8. Do any of the time series examples above look like white noise? No. White noise is not a great model for time series data, which typically has both trends (nonconstant mean and variance) and dependence (nonzero correlation). But it is an important concept and will serve as a building block for more complex models

# 5 Linear filtering

- Another important concept in time series is *filtering*. fields. A *linear filter* is just the result of performing a moving linear combination of a series $x_t$, $t = 1, 2, 3, \ldots$, with given weights. (Nonlinear filters take nonlinear combinations and we won't talk about them)
- The simplest and most common type of linear filter is a moving average. For example, a *moving average*, that is centered around lag 0, of window length 3, is

$$y_t = \frac{1}{3}\Big(x_{t-1} + x_t + x_{t+1}\Big)$$

  In principle, we could center the moving average wherever we want. But the term moving average (without further specification) usually means that we center it at lag 0

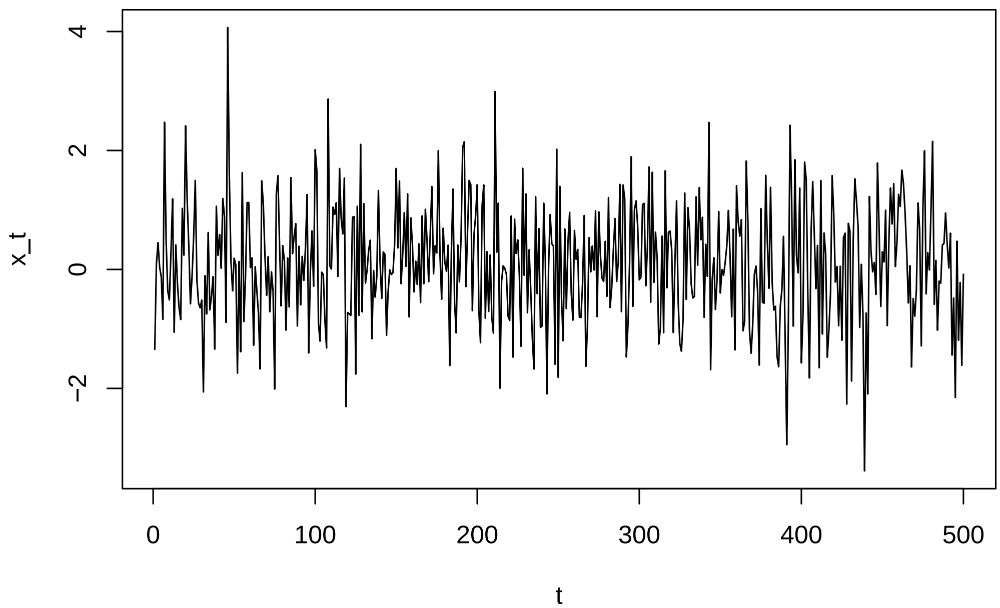


Figure 8: *Gaussian white noise.*

- When we center the moving average so that its right endpoint is at time $t$, ensuring we only average past values, this is called a *trailing average*. For example, a trailing average of length 3 is

$$y_t = \frac{1}{3}\Big(x_{t-2} + x_{t-1} + x_t\Big)$$

  The Covid-19 data plotted above (Figures 6 and 7) was actually filtered with a 7-day trailing average

- A general linear filter takes the form

$$y_t = \sum_{i=-\infty}^{\infty} a_i x_{t-i}$$

  for constants $a_i$, where typically only finitely many are nonzero. For example, the second to last example took $a_{-1} = a_0 = a_1 = 1/3$, and the last example took $a_0 = a_1 = a_2 = 1/3$

- Linear filters provide a form of *smoothing* for time series, which we'll revisit briefly later in the lecture, and then in more detail in a future week

# 6 Autoregression

- An *autoregressive process* is one that takes the form

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + w_t$$

  for coefficients $\phi_1, \dots, \phi_p$ and errors $w_t$

- Typically we assume that the errors are a white noise sequence

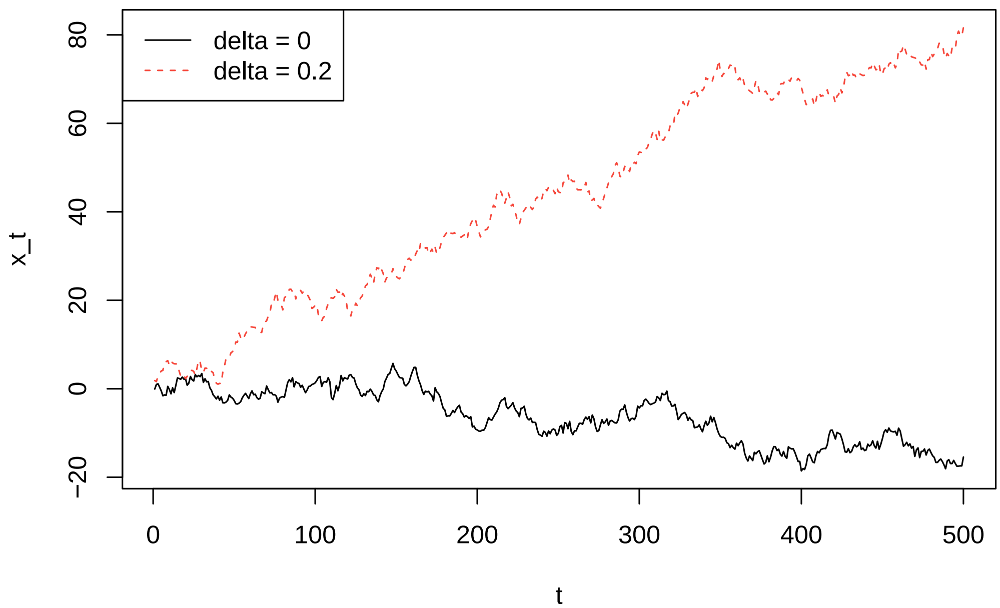


Figure 9: *Random walk without and with drift.*

- The value of $p$ is called the *order* of the autoregressive process, and the abbreviation AR($p$) is common. So, for example, AR(3) means an autoregressive process with lag 3: each value in the sequence depends on the last 3 values
- The simplest autoregressive process is AR(1), with coefficient $\phi_1 = 1$: this is

$$x_t = x_{t-1} + w_t$$

  also known as a *random walk*
- Random walks may seem very simple and trivial at first but actually they and simple generalizations are pretty fascinating, and important
- For example, did you know that a random walk in 1 and 2 dimensions is *recurrent* (returns to where it started—say, the origin—infinitely often with probability 1), but in 3 dimensions and higher it is *transient* (returns to the origin infinitely often with probability 0)
- (And did you know that Larry Brown proved in 1971[1] that this last fact is *equivalent* in some precise sense to Stein's paradox: that the MLE in a normal means model is admissible in dimensions 1 and 2, and inadmissible in dimensions 3 and higher??)
- You could also say that random walks were the beginning of what made Google the giant they are today (the "billion dollar eigenvector")
- Ok, back to the main story, a *random walk with drift* takes the form

$$x_t = \delta + x_{t-1} + w_t$$

  for some $\delta > 0$. Figure 9 plots examples of random walks with and without drift

[1] Larry Brown (1971), "Admissible Estimators, Recurrent Diffusions, and Insoluble Boundary Value Problems"

- By unraveling the last iteration, we can write a random walk with drift equivalently as (assuming we start at $x_0 = 0$):

$$x_t = \delta t + \sum_{i=1}^{t} w_i$$

Think: what happens to the mean and variance of this as $t$ grows?

# 7 Signal plus noise

- A useful general time series model is called the *signal plus noise model*, of the form

$$x_t = \theta_t + w_t$$

where the errors $e_t$, $t = 1, 2, 3, \ldots$ may be white noise or may be correlated over time

- The problem of estimating the signal $\theta_t$, $t = 1, 2, 3, \ldots$ is of great interest in many applications
- It is common in time series to think about *decompositions* for the signal sequence, into a *trend* $u_t$ and *seasonal* components $s_t$:

$$\theta_t = u_t + s_t$$

- The seasonal component $s_t$ has a regular/periodic behavior for some fixed period. For example:
  - Routine doctor's office visits dip on weekends (weekly period)
  - Gambling goes up at the beginning of each month (monthly period)
  - Chocolate purchases go up on and around Valentine's day (yearly period)
- The trend component $u_t$ is not regular, and is typically not assumed to be linear or to have any particular parametric form; it is typically estimated *nonparametrically* using some kind of smoother—more on this later in the course
- (Some authors even further decompose the trend into two components: *proper trend* and *cycle*. The former is monotone and the latter has a cyclic behavior but without a fixed period. We don't generally find this a useful distinction and won't really pursue this ... but it may be good to know about in case you hear people, particularly economists, mentioning: trend, seasonal, and cyclic components separately)
- Economists and official statistics agencies (like the US Census Bureau) care a lot about decompositions into trend and seasonal components ... there are various methods for doing so that we may cover later in the course: what is considered the "classical" decomposition, but also X-11 (developed by the US Census Bureau and Statistics Canada), SEATS (developed by the Bank of Spain), and STL (developed by academics at the University of Michigan and Bell Labs)
- Many consider STL to be the most general and robust method for decomposition. Figure 10 gives an example applied to US retail employment data

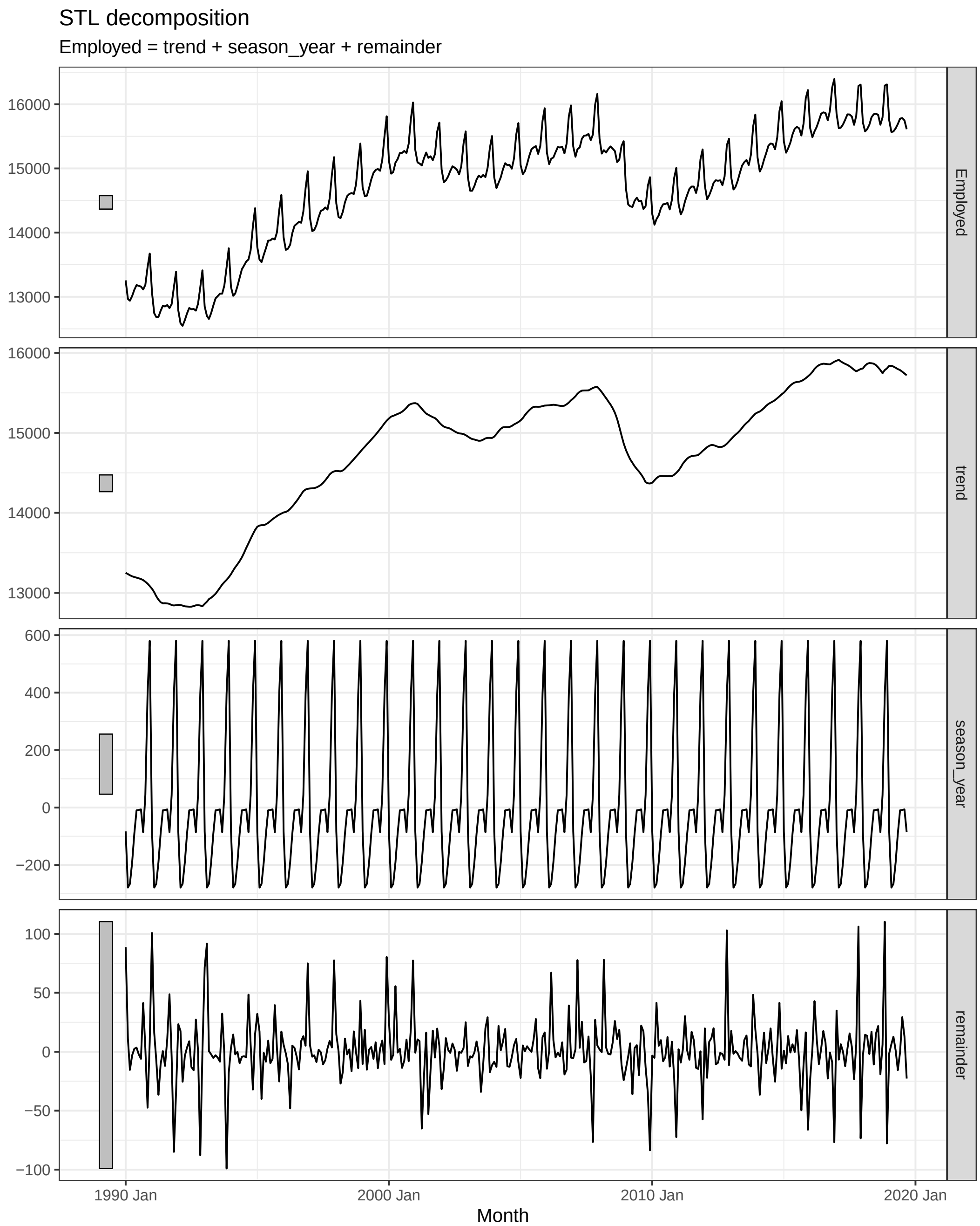


Figure 10: *STL decomposition of US retail employment data (from HA).*