Lecture 2a - Seasonality

Jared Fisher

Lecture 2a

Announcements

- Homework 1 is due TOMORROW, Wednesday Feb 3, by 11:59pm via Gradescope, and IS graded.
- Homework 1 covers material from weeks 0 and 1 (everything covered before this week): noise, stationarity, trends.
- Provost's message: finals week conflicts with multiple religious holidays. If you have a schedule conflict of any kind, "we ask that students request accommodations no later than Friday, February 12, 2021."
- Project Checkpoint 1 is due next week. More details on Thursday!

Recap

Our Modeling Approach

- For time series Y_t:
$$Y_t = signal(t) + X_t$$
- Where X_t is a **stationary process**
- So we'll model $signal(t)$ such that $y_t - \widehat{signal}(t)$ looks stationary
- ASIDE: The book's main variable is x_t, whether data or stationary process. In this class I choose to clarify that by having y_t be data and X_t be a stationary process.

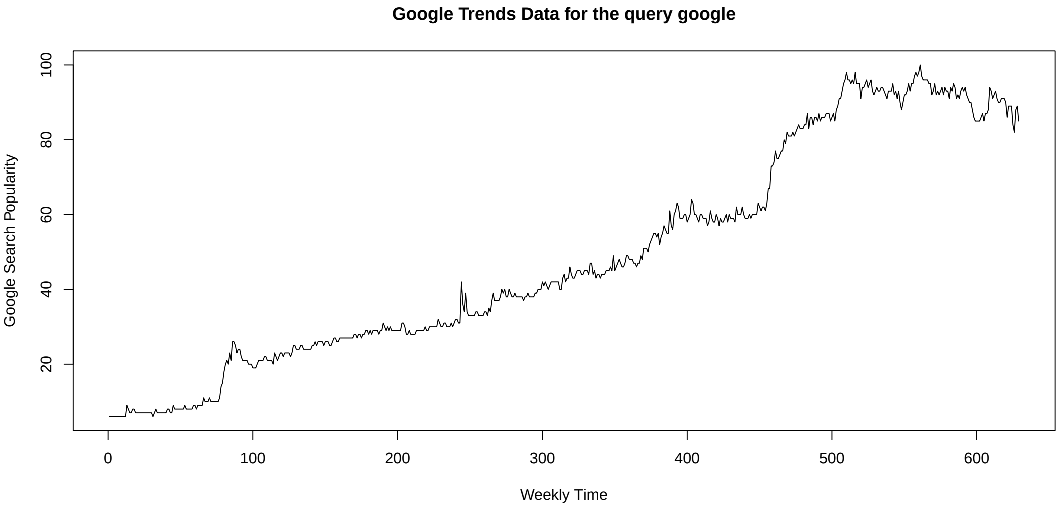
Trend Models

- If there are no seasonal effects, $Y_t = m_t + X_t$
- m_t is the trend
- X_t is a stationary process, perhaps white noise
- **Idea:** Remove trend, so that the residuals exhibit steady behavior over time, i.e. looks stationary. Stationarity gives us a structure we can use to create models, predictions/forecasts, etc.

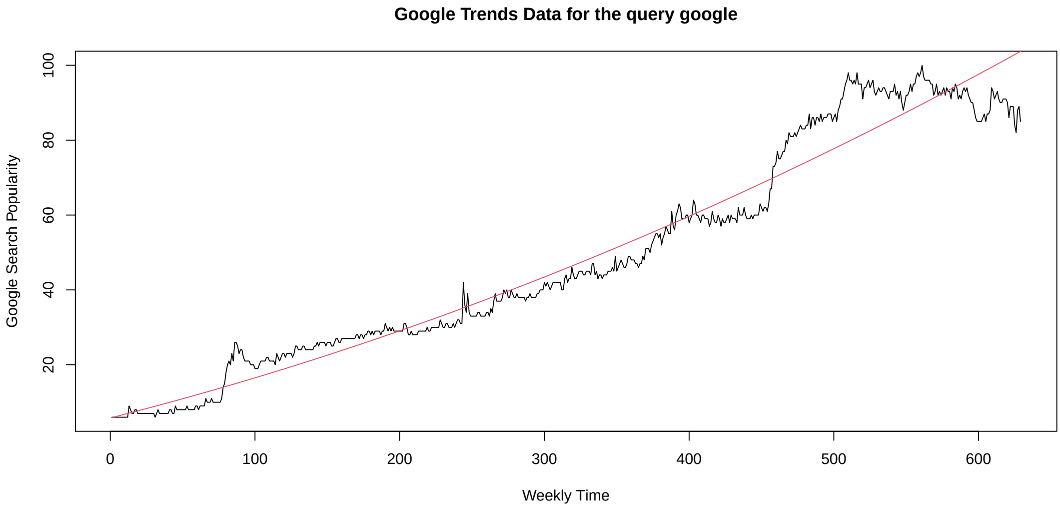
Trend Estimates

- Parametric form e.g. quadratic: $\hat{m}_t = \alpha + \beta t + \gamma t^2$

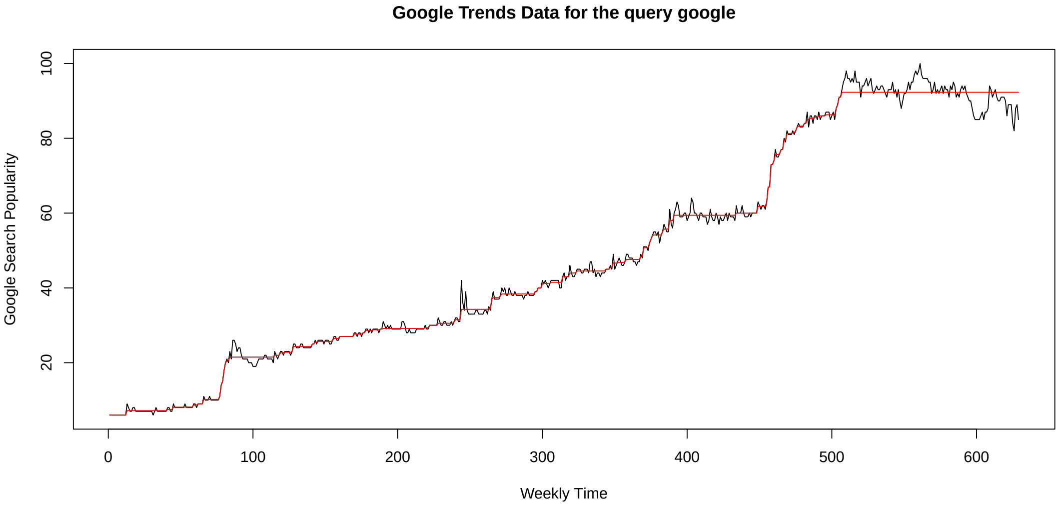
Example - Googling “Google”



Google + Parametric



Google + Isotonic



\# New Material

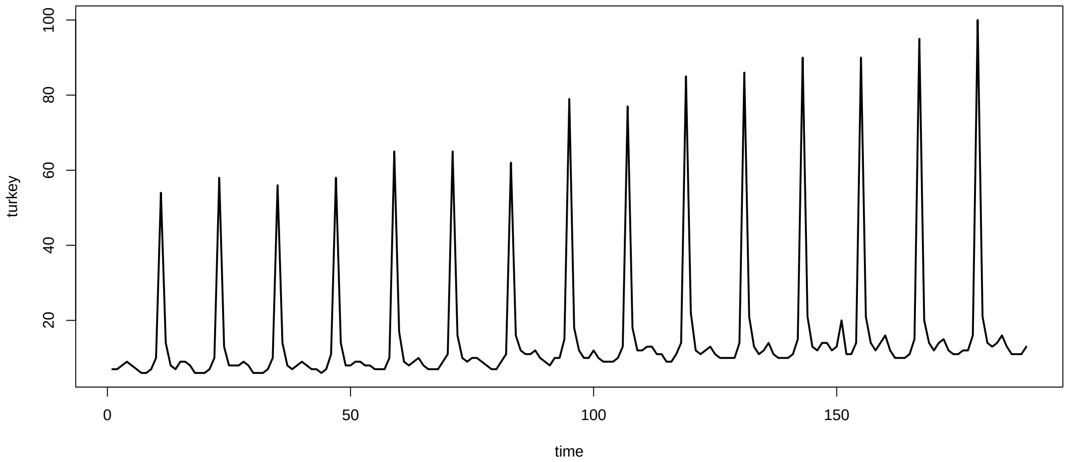
Today's Assumptions

$$Y_t = m_t + s_t + X_t$$

- m_t is the **deterministic** trend
- s_t is the **deterministic** seasonal effect
- X_t is a stationary process, perhaps white noise
- **Idea:** Remove trend and seasonality, so that the residuals exhibit steady behavior over time, i.e. looks stationary.

Seasonality

Example: Google Searches for “Turkey”

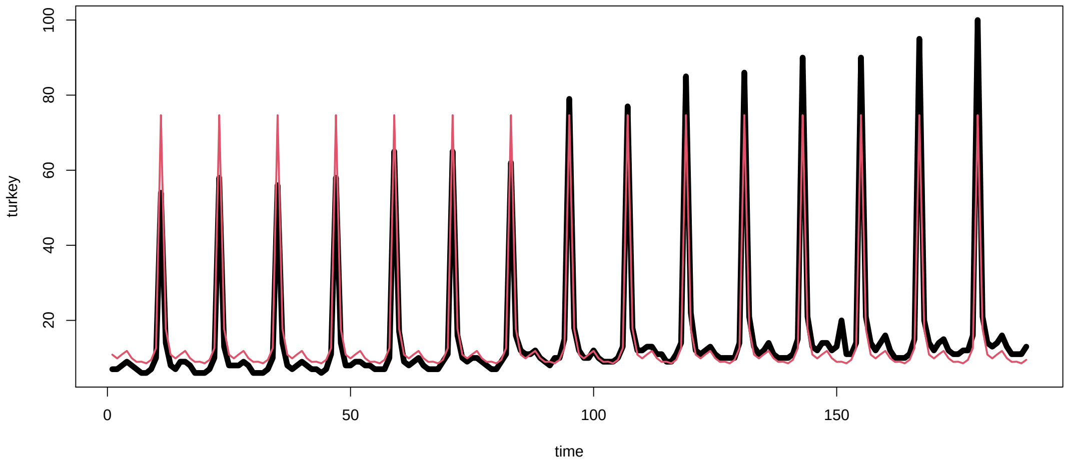


Super-simple Seasonality Model

$$Y_t = s_t + X_t$$

- ALWAYS: s_t is a periodic function of a known period d such that $s_{t+d} = s_t$
- X_t is white noise
- So, what should be the period for our turkey example (monthly data)?

Turkey with $d = 12$ via Indicator Variables



“Removing” the Seasonal Component

If we can “remove” the seasonal component, we can check the residuals!

Our main two methods for estimating s_t:

1. Parametric form (linear model)
2. Nonparametric seasonality estimation (remove noise by averaging)

Parametric Seasonality Function

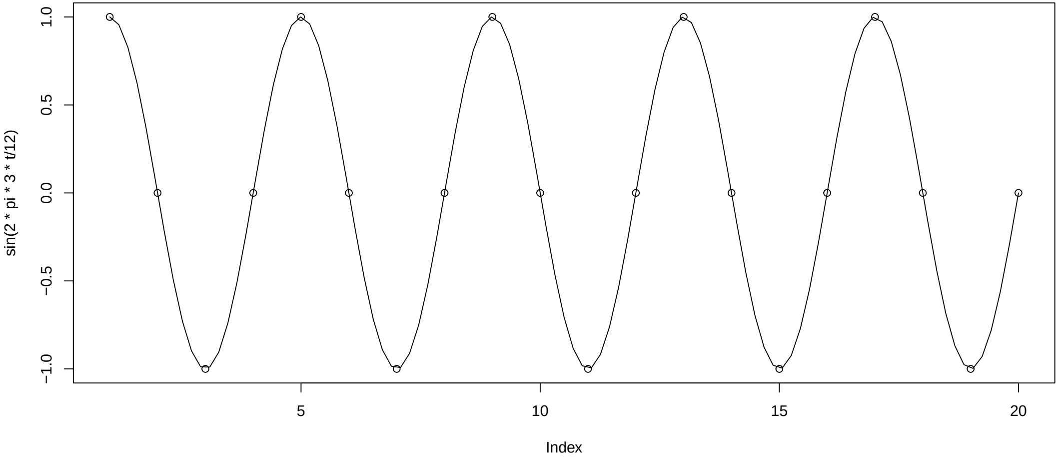
The parametric seasonality function for seasonality with known period d (and discrete, uniformly sampled time points)

$$s_t = \sum_{k=1}^{K} \left(a_k \cos(2\pi tk/d) + b_k \sin(2\pi tk/d) \right)$$

- a is the “Amplitude”
- $f = k/d$ is the “Frequency”
- d/k is the “Period”
- No need for $K > d/2$

Parametric Seasonality Function

Example: $sin(2\pi tk/d)$ with $d = 12$ and $k = 3$



Nonparametric Seasonality

$$\hat{s}_i := \text{average of } X_i,\, X_{i+d},\, X_{i+2d},\, \ldots$$

Note though that we're fitting d parameters with n observations. n must be sufficiently larger than d.

Trend and Seasonality Models

Trend and Seasonality

$$Y_t = m_t + s_t + X_t$$

- m_t is the trend (e.g., approximately linear or quadratic)
- s_t is the periodic function of know period d, $s_{t+d} = s_t$
- X_t is a stationary process (e.g. white noise)

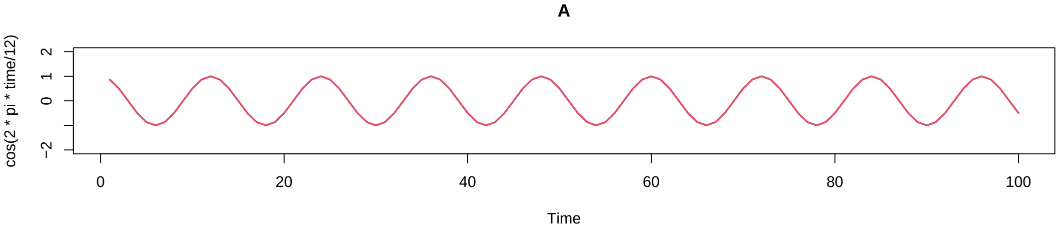
Idea: Remove both trend and seasonality so that the residuals exhibit steady behavior over time

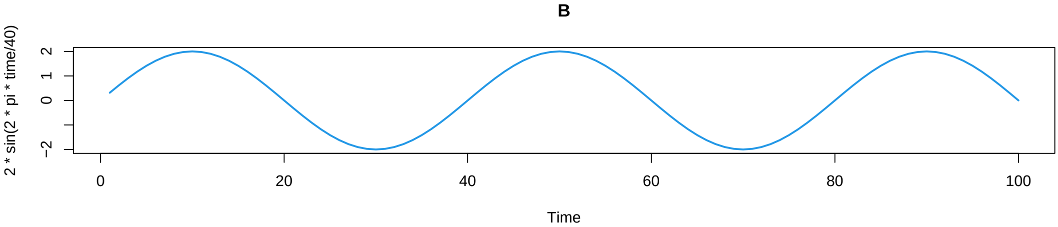
Code

Let's check out some code for all this!

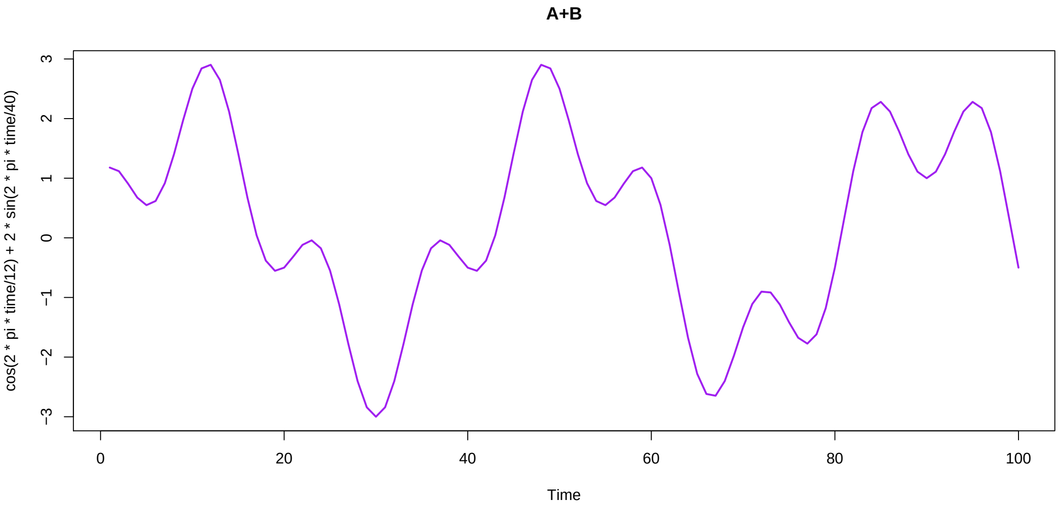
Frequency Domain

Example: Seasonality





Example: Seasonality



Recall: Parametric Seasonality Function

$$s_t = \sum_{k=1}^{K} \left(a_k \cos(2\pi tk/d) + b_k \sin(2\pi tk/d)\right)$$

- But how do we wisely choose the frequency k/d to include? In other words, do we need all K?
- What if there is no clear value of d?

Transition to Frequency domain

- This class is largely about the time domain approach: models constructed via the relationship of observations Y_t at different time points.
- Sometimes though, we will look at a time series as a composition of periodic components with different frequencies.
- This is quite natural for many time series data, which are often directly driven by periodic random events, like the purple curve in the "A+B" example a couple slides ago.

Definition: Sinusoids

We define the set of sinusoid functions as

$$\{g(t) = R\cos(2\pi ft + \Phi) \;:\; R \in R_+,\; f \in R_+,\; \Phi \in [0, 2\pi/f)\},$$

where

- R is called the *amplitude*
- f is called the *frequency*
- Φ is called the *phase*
- $1/f$ is called the *period*

Demonstrations in R

Let's explore these parameters to get a sense of what they imply

Sinusoids rewritten a different way

- Estimating the phase shift Φ is nontrivial with the tools in this class, but we can rewrite the sinusoid equation to be more convenient (you will show this in lab).
- With $A = R\cos(\Phi)$ and $B = -R\sin(\Phi)$ one can rewrite sinuosoids as

$$\{g(t) = A\cos(2\pi ft) + B\sin(2\pi ft) \ : \ A, B \in R, \ f \in R_+\}.$$

- This is helpful as we can find the coefficients A and B with linear models, but that means we must find the appropriate frequencies f first. The frequency domain will help with this!

Reading Alignment:

- Decomposing the time series into trend/seasonality/noise is only addressed in spots in this textbook...
- The intro to Chapter 4 and the first part of Section 4.1 talk about seasonality (with different words) and begin the discussion of sinusoids and the frequency domain.