Stationarity

Ruoqi Yu

Week 2

Section 1

Announcements

Assignments

- The Practice Homework (sometimes called HW0) is “due” today (9/1), but you can submit it anytime for practice. The solution will be posted this week.
- Your first graded assignment is the Homework 1, which was posted on bCourses, and is due before 3 pm, September 8, on bCourses.

Clarification on the Syllabus

- Project: Form your own group of 3-5 classmates
- Midterm: No make-up midterm. If you can't take the midterm exam in person, the weight will be shifted to the final.
- Final: If you can't take the final exam in person, you will fail this course.
- Homework: You will get approximately one week for each homework.

Waitlist and Concurrent Students

- I can add anyone on the waitlist/CE to bCourses so you can access materials. Please send me an email if you can't access bCourses.
- Being on bCourses does not mean you are enrolled in the course.
- I will begin processing applications when waitlist has emptied, and the stat office will give final approval after that.
- I will accept applications in the order they were received.

Section 2

Recap

Time Series

- (Univariate time series) we don't have n subjects at 1 point in time, we have 1 subject at n points in time...
- Our model: $Y_t = f(t) + X_t$
- $f(t)$ can be thought of as the signal
- X_t can be thought of as the noise

Definitions (TSA4e Example 1.8)

Random variables $X_1, ..., X_n$ will be denoted as

- White noise: if they have mean zero, variance σ^2, and are uncorrelated
- IID noise: if they are white noise AND are independent and identically distributed (IID).
- Gaussian [white] noise: if they are IID noise AND are normally distributed, $X_i \sim N(0, \sigma^2)$

Definitions:

- Autocovariance (Definition 1.2):

$$\begin{aligned}\gamma_x(s,t) &= Cov(X_s, X_t)\\ &= E\left[(X_s - E[X_s])(X_t - E[X_t])\right]\end{aligned}$$

- [mental aside: let $s > t$ and $h = s - t$. h is the number of "lags"]
- Sample autocovariance (Definition 1.14):

$$\hat{\gamma}(h) = n^{-1}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})$$

Definitions:

- Autocorrelation function "ACF" (Definition 1.3):

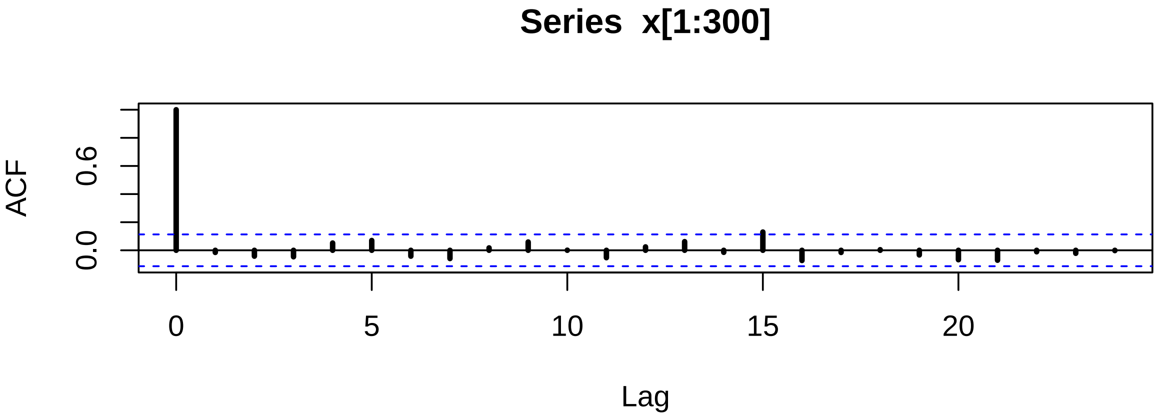
$$\rho(s,t) = \frac{\gamma_x(s,t)}{\sqrt{\gamma_x(s,s)\gamma_x(t,t)}} = \frac{Cov(X_s, X_t)}{\sqrt{Var(X_s)Var(X_t)}}$$

- Sample autocorrelation (Definition 1.15):

$$r_h = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

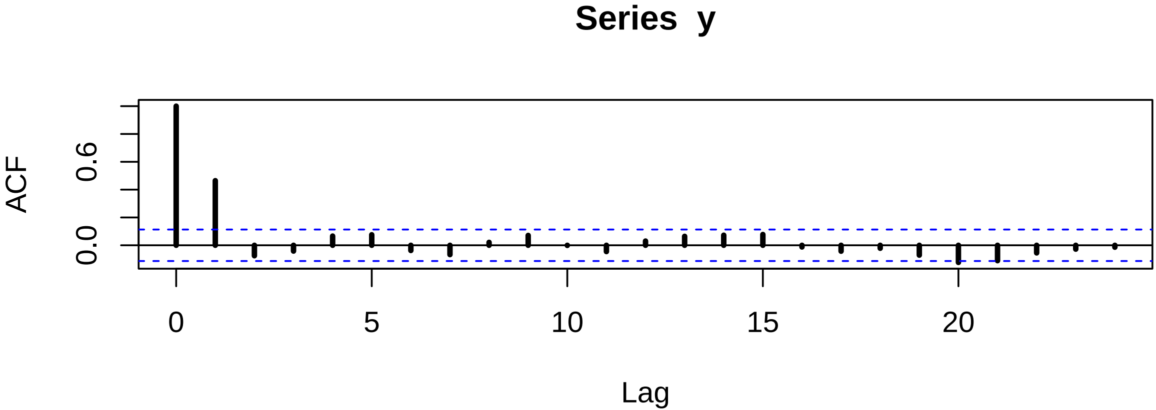
ACF plot/Correlogram

```
x = rnorm(301)
acf(x[1:300],lwd=3)
```



ACF plot - simple Moving Average

```
y = .5*(x[1:300] + x[2:301])
acf(y,lwd=3)
```



CI for Sample Correlations

- Wouldn't it be great if those dashed blue lines were the appropriate confidence interval?
- A value of r_h outside the blue bands is significant, i.e., it gives evidence against pure white noise.

Simplified Theorem A.7 (see Property 1.2)

- Under general conditions, if x_t is white noise, then for any fixed H, the sample autocorrelations

$$\sqrt{n}\begin{pmatrix} r_1 \\ \vdots \\ r_H \end{pmatrix} \to N(0, I) \quad \text{as } n \to \infty$$

- Key takeaway: $var(r_h) = 1/n$ (Equation 1.38)

Confidence Interval

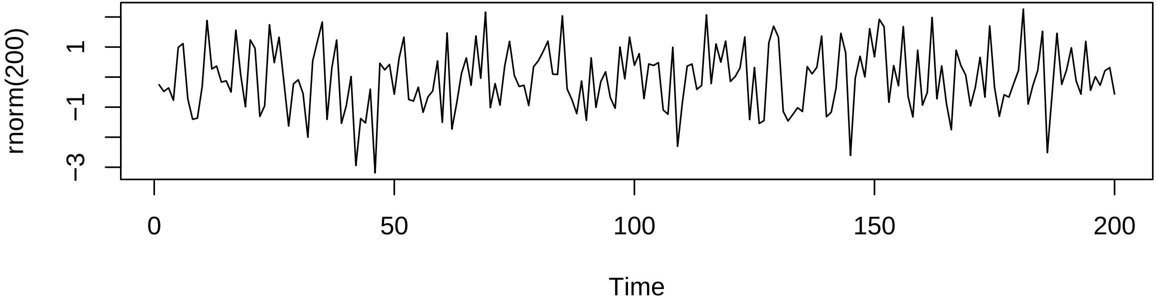
- What is the 95% CI for r_h?
- For white noise $r_1, ..., r_H \overset{iid}{\sim} N(0, 1/n)$, for

$$P\left(|r_h| > 1.96n^{-\frac{1}{2}}\right) \approx P\left(|N(0,1)| > 1.96\right) = 5\%$$

- So for $n = 100$, $1.96n^{-\frac{1}{2}} = 1.96/\sqrt{100} = .196$
- Note. The overall probability of getting at least one r_h outside the bands increases with the number of coefficients plotted.
  - If 20 r_h's are plotted, we expect to get 1 $(= 20 \times 5\%)$ significant value under pure white noise.

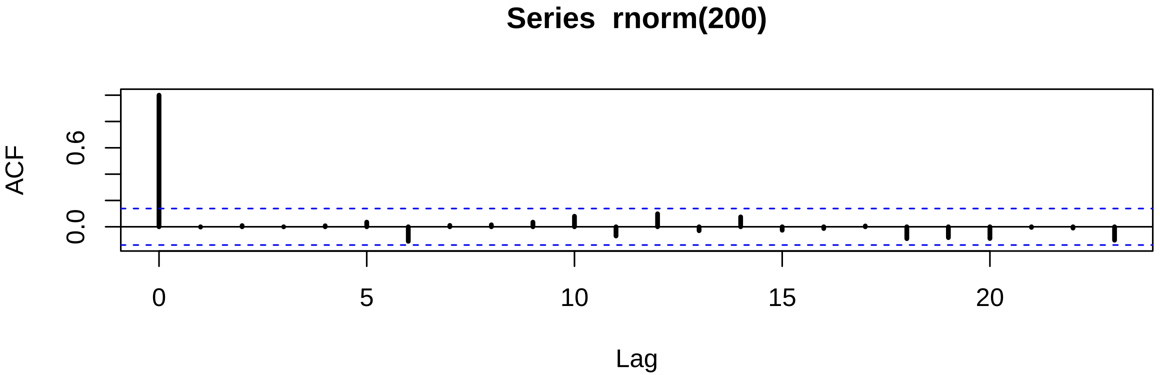
Gaussian Noise

```
plot.ts(rnorm(200))
```



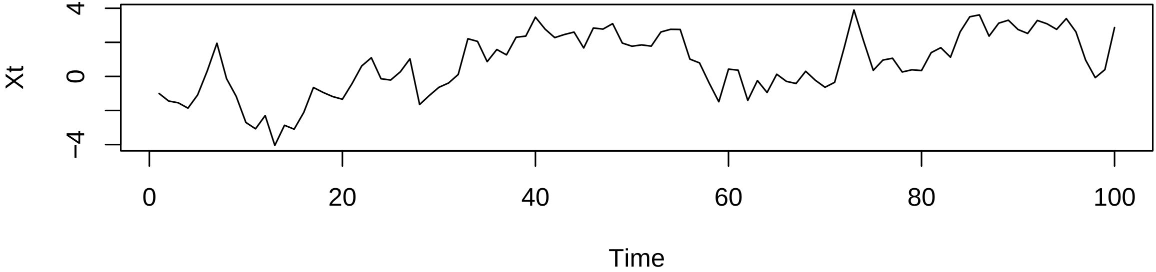
ACF plot - Dashes at .196?

```
acf(rnorm(200),lwd=3)
```



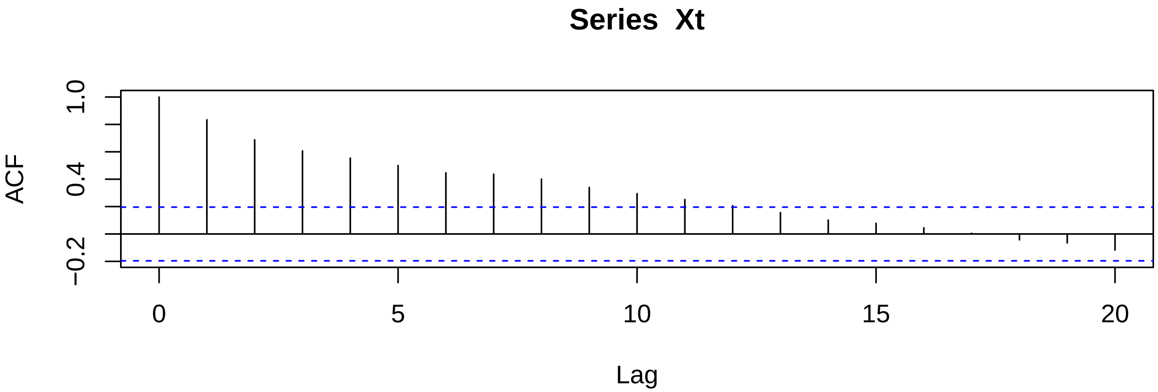
Autoregressive Process AR(1) (we'll learn about it later)

```
Xt = arima.sim(model=list(ar=.9),n=100)
plot.ts(Xt)
```



ACF Plot of AR(1)

```
acf(Xt)
```



Section 3

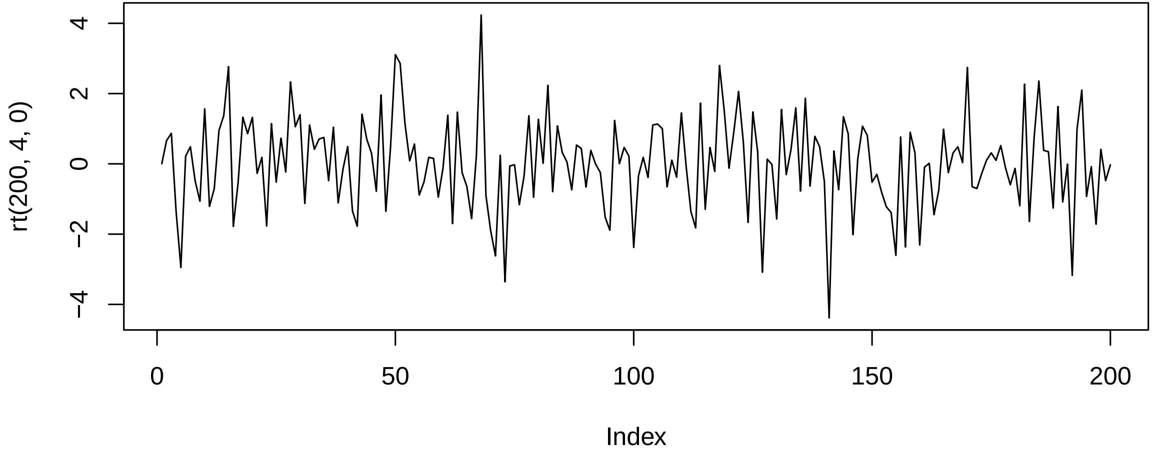
Expand our definition of Noise

Modeling Idea

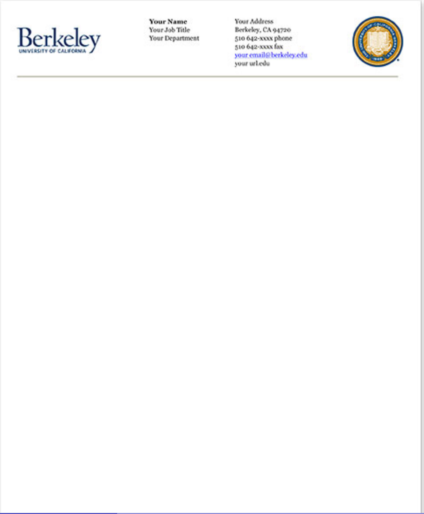
- $Y_t = f(t) + X_t$
- In this class we'll first model f such that $y_t - \hat{f}(t)$ exhibits steady behavior over time
- But what do we mean by "steady" or "stable"?
- White noise is correct, but we're going to look at a broader group of processes as steady
- "Stationary" is the adjective, "Stationarity" is the noun

White Noise

```
plot(rt(200,4,0),type='l')
```



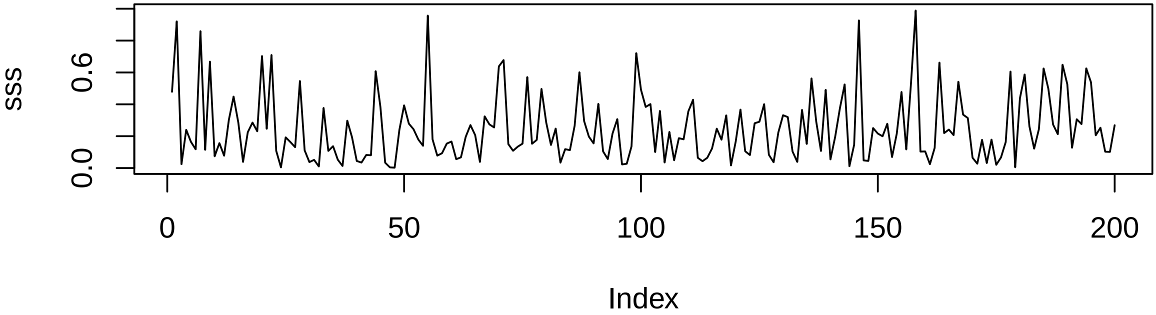
Stationery


Berkeley
UNIVERSITY OF CALIFORNIA
Your Name
Your Job Title
Your Department
Your Address
Berkeley, CA 94720
510 642-xxxx phone
510 642-xxxx fax
your email@berkeley.edu
your url.edu

Stationary (but not White Noise, why?)

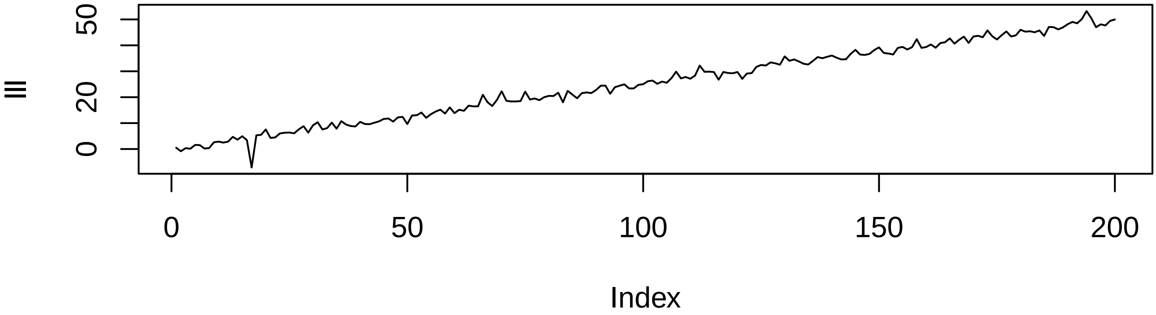
```
sss = rexp(200,4)
plot(sss,type='l')
```

E not equals to 0



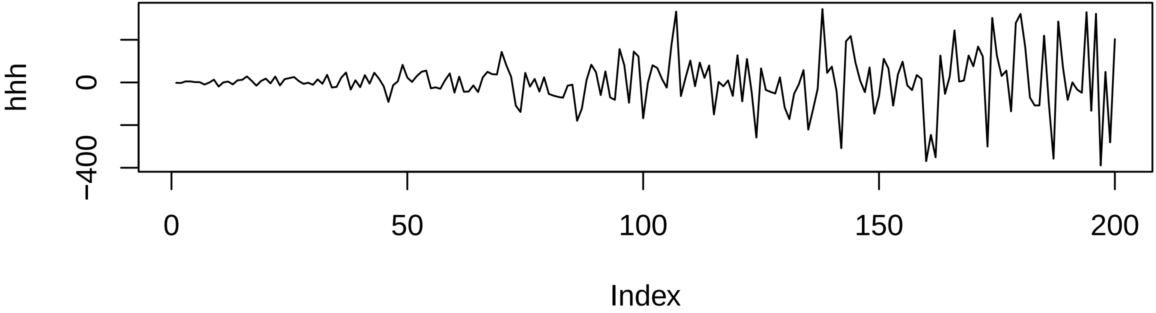
Not Stationary - linear trend

```
lll = (1:200)/4+rt(200,4,0)
plot(lll,type='l')
```

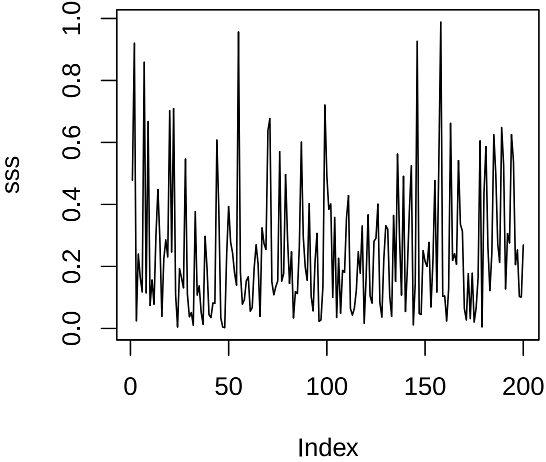


Not Stationary - heteroskedastic

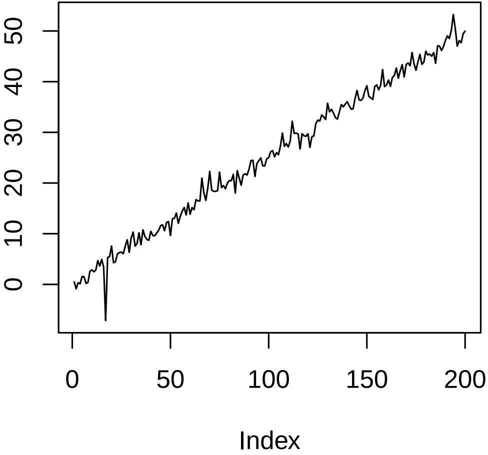
```r
hhh = rnorm(200,0,1:200)
plot(hhh,type='l')
```



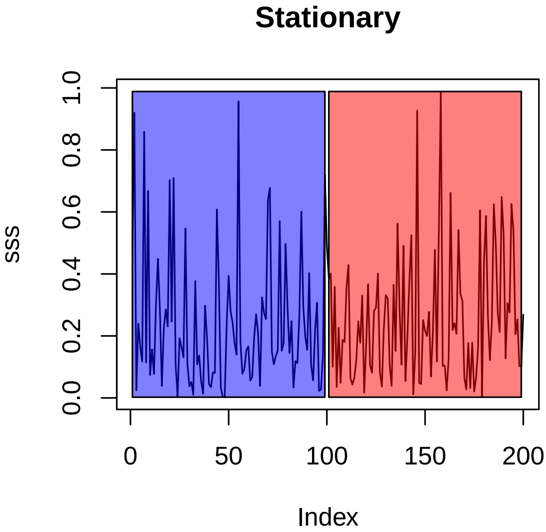
Side by Side


sss
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
Index

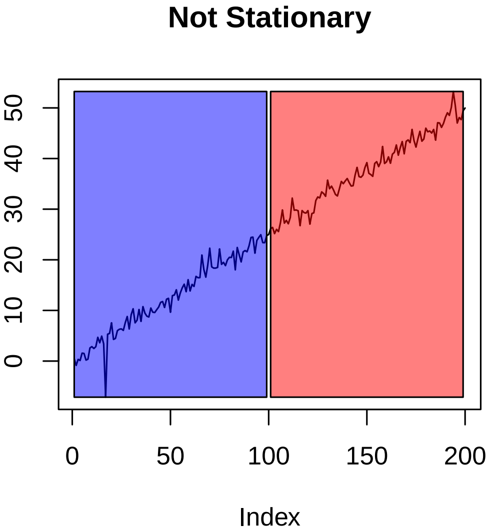
≡


0
10
20
30
40
50
0
50
100
150
200
Index

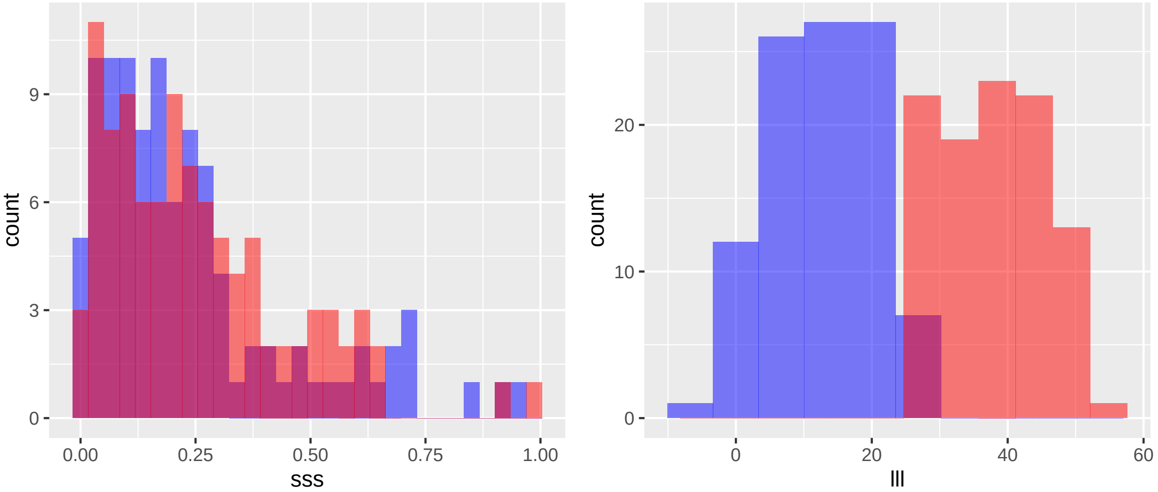
Side by Side


Stationary
sss
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
Index

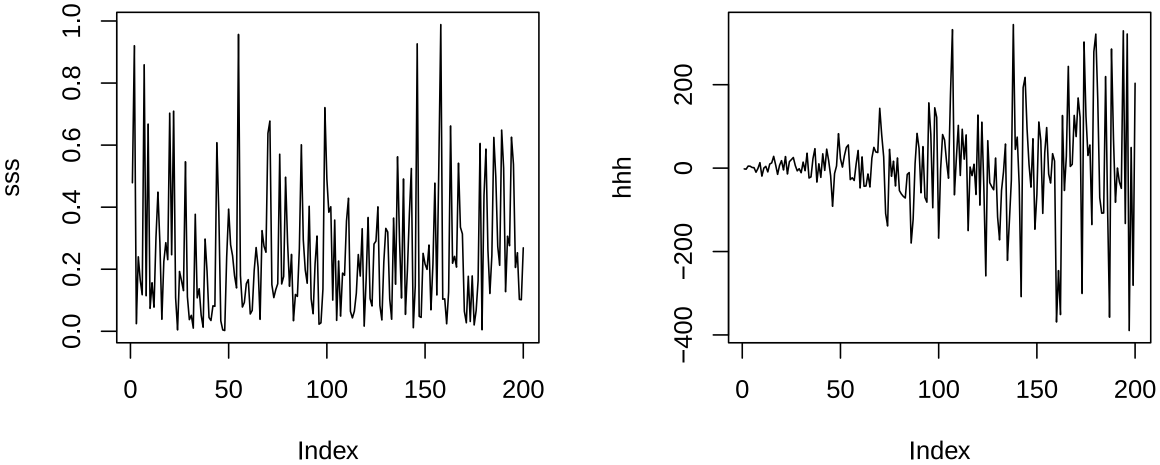
≡


Not Stationary
0
10
20
30
40
50
0
50
100
150
200
Index

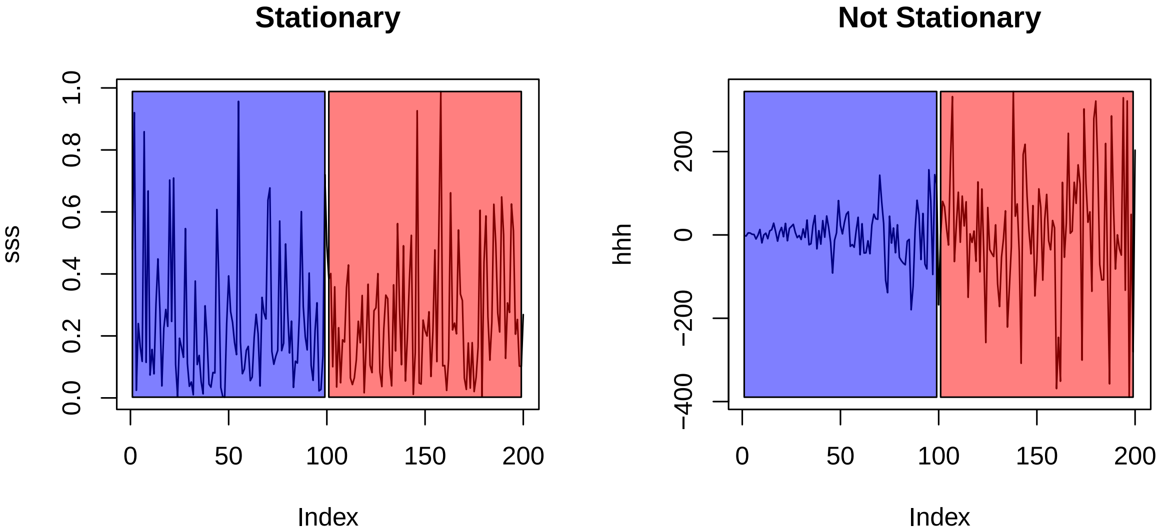
Side by Side



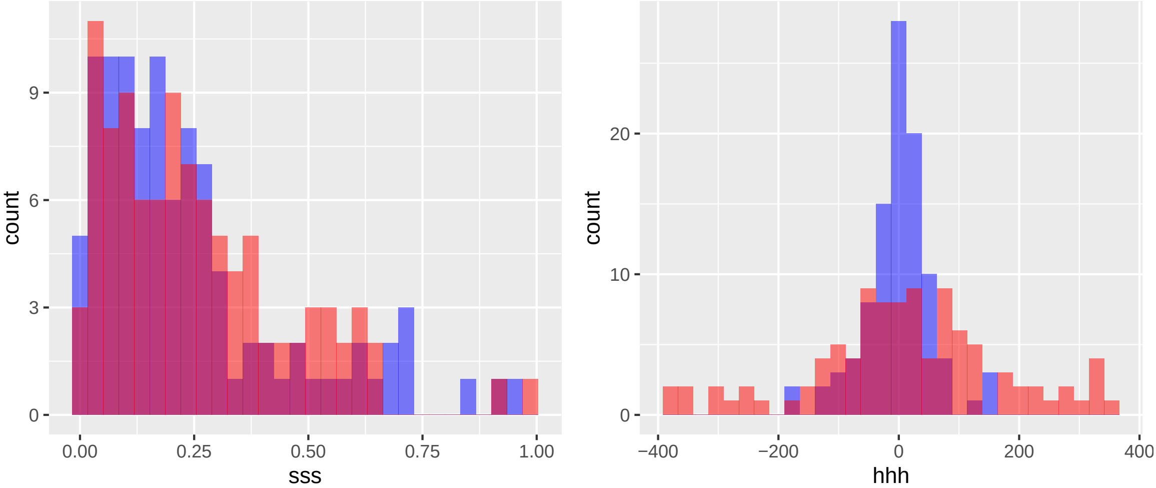
Side by Side


sss
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
Index
hhh
−400
−200
0
200
0
50
100
150
200
Index

Side by Side


Stationary
sss
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
Index
Not Stationary
hhh
−400
−200
0
200
0
50
100
150
200
Index

Side by Side


count
9
6
3
0
0.00
0.25
0.50
0.75
1.00
sss
count
20
10
0
−400
−200
0
200
400
hhh

Stationary

Stationary essentially means that the dependence structure of $\{X_t\}$ is invariant over time and hence, we can learn while observing more and more data.

Strongly/Strictly Stationary (TSA4e Definition 1.6)

A strictly stationary time series is one for which the probabilistic behavior of every collection of values $(X_{t_1}, X_{t_2}, ..., X_{t_k})$ is identical to that of the time shifted set $(X_{t_{1+h}}, X_{t_{2+h}}, ..., X_{t_{k+h}})$. That is,

$Pr(X_{t_1} \leq c_1, ..., X_{t_k} \leq c_k) = Pr(X_{t_{1+h}} \leq c_1, ..., X_{t_{k+h}} \leq c_k)$ for

- all $k = 1, 2, ...$
- all time points $t_1, t_2, ..., t_k$
- all numbers $c_1, c_2, ..., c_k$
- all time shifts $h = ..., -2, -1, 0, 1, 2, ...$

Strictly Stationary - Rephrased Definition

A doubly infinite sequence of random variables {Xt}:

$$..., X_{-2}, X_{-1}, X_0, X_1, X_2, ...$$

is strictly stationary if for every choice of times $t_1, ..., t_k$ and lag h, the joint distribution of

$$(X_{t_1}, X_{t_2}, ..., X_{t_k})$$

is the same as the joint distribution of

$$(X_{t_{1+h}}, X_{t_{2+h}}, ..., X_{t_{k+h}}).$$

Merits and challenge

- Stationarity means that the joint distribution of the random variables remains constant over time. For example, under stationarity, the joint distribution of today's and tomorrow's random variables is the same as the joint distribution of the variables from any two successive days (past or future).
- Note how stationarity makes the problem of forecasting or prediction feasible. From the data, we can study how a particular day's observation depends on the those of the previous days and because under stationarity, such a dependence is assumed to be constant over time, one can hope to use it to predict future observations from the current data.
- Really hard, if not impossible, to show a time series is strictly stationary in applied practice.
- So we'll usually look at something else.

Weakly Stationary (TSA4e Definition 1.7)

A weakly stationary time series, X_t, is a finite variance process such that

1. the mean value function, $E(X_t)$, is constant and does not depend on time t, and
2. the autocovariance function, $\gamma(s, t)$ depends on s and t only through their difference $|s - t|$.

Weakly Stationary - Rephrased Definition

A doubly infinite sequence of random variables $\{X_t\}$ with finite variance is weak stationary if

1. The mean of the random variable X_t, denoted by $E(X_t)$, is the same for all times t
2. The covariance between X_t and X_s is the same as the covariance between X_{t+h} and X_{s+h} for every choice of times t, s, and lag h.

Weakly Stationary

- Weak stationarity means that the second order properties (means and covariances) of the random variables remain constant over time. Unlike strong stationarity, the joint distribution of the random variables may well change over time.
- Another way of phrasing the condition in 2: autocovariance $\gamma(s, t) = Cov(X_s, X_t)$ and autocorrelation $\rho(s, t) = \gamma(s, t)/\sqrt{Var(X_s)Var(X_t)}$ only depend on the time lag $|t - s|$ between them.

Weakly Stationary

- For weakly stationary sequences {X_t} we define the autocovariance function (ACVF)

$$\gamma(h) = \gamma(t, t+h) = Cov(X_t, X_{t+h})$$

and the autocorrelation function (ACF)

$$\rho(h) = \gamma(h)/\gamma(0).$$

- Property:

$$\gamma(h) = \gamma(-h), \rho(h) = \rho(-h).$$

Stationarity

In this course, when we say “stationary” we will always mean “weakly stationary”.

Example: White Noise

Let $\{X_t\}$ be white noise. Recall the definition of white noise:

- $E(X_t) = 0, \quad \forall t$
- $Var(X_t) = \sigma^2, \quad \forall t$
- $Cov(X_t, X_s) = \begin{cases} 0 & t \neq s \\ \sigma^2 & t = s \end{cases}$

Is White Noise stationary?

Example: White Noise

Let $\{X_t\}$ be white noise. Recall the definition of white noise:

- $E(X_t) = 0, \quad \forall t$
- $Var(X_t) = \sigma^2, \quad \forall t$
- $Cov(X_t, X_s) = \begin{cases} 0 & t \neq s \\ \sigma^2 & t = s \end{cases}$

Is White Noise stationary?

Example: White Noise

- $E(X_t)$ is the same for all t
- $Cov(X_t, X_s)$ only depends on $|s-t|$ not on s or t individually:
- $|s-t| = 0 \Rightarrow Cov(X_t, X_s) = \sigma^2$
- $|s-t| > 0 \Rightarrow Cov(X_t, X_s) = 0$
- ACVF:

$$\gamma(h) = \begin{cases} 0 \text{ if } h \neq 0 \\ \sigma^2 \text{ if } h = 0 \end{cases}$$

- ACF:

$$\rho(h) = \begin{cases} 0 \text{ if } h \neq 0 \\ 1 \text{ if } h = 0 \end{cases}$$

Example: White Noise

- Yes, white noise is stationary, and we similarly see that IID noise and Gaussian noise are stationary too.
- White noise is only a very special example of a stationary time series. Stationarity allows for considerable dependence between successive random variables in the series. The only requirement is that the dependence should be constant over time.

Example: IID Noise

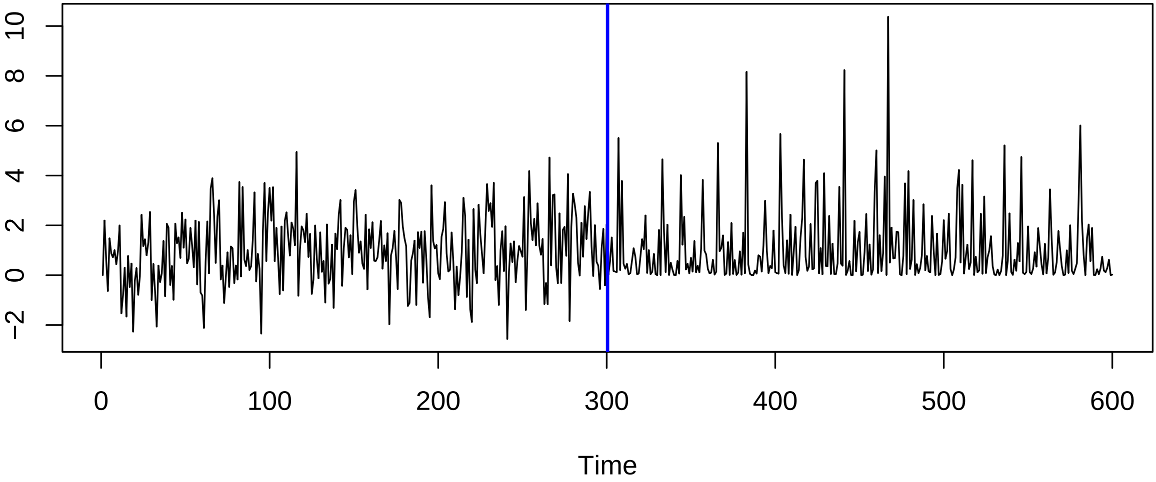
Let $\{X_t\}$ be IID noise. Recall the definition of IID noise:

- $\{X_t\}$ is white noise
- X_t and X_s are independent for all $t \neq s$
- $X_t \sim A$ for all t (identically distributed)

Is IID Noise strongly stationary?

- For $k = 1$ and any h, the joint distribution of $(X_{t_1}, X_{t_2}, ..., X_{t_k})$ is the same as the joint distribution of $(X_{t_{1+h}}, X_{t_{2+h}}, ..., X_{t_{k+h}})$ for $k = 1$ (3rd point above)
- For $k > 1$, because of independence, the multivariate distribution will be the same too.
- So yes, IID noise is strongly stationary, as is Gaussian noise (The case where A is the Normal/Gaussian distribution).

Example: Half Gaussian (IID $N(1,4)$), Half Chi-Square (IID $\chi^2(1)$)



Exercise

- If $X_t = \frac{1}{3}(W_t + W_{t-1} + W_{t-2})$ with $W_t \sim WN(0, \sigma^2)$, is X_t stationary?
- If $Y_t = \mu + Y_{t-1} + W_t$ with $W_t \sim WN(0, \sigma^2)$ and $Y_0 = 0$, is Y_t stationary when $\mu = 0$?
- If $Z_t = \alpha + \beta t + W_t$, where W_t is stationary. Is Z_t stationary? What is the ACVF of Z_t?

A Few Extra Notes on Stationarity

- Note that the concept of stationarity (both weak and strong), as well as the notion of autocovariance and autocorrelation functions $\gamma(h)$ and $\rho(h)$ applies to the random variables $\{X_t\}$, not to a specific data set (which is a single realization of the random variables $\{X_t\}$).
- Thus, strictly speaking, we cannot say that a particular data set is stationary. We can only say that a particular data set is a realization of stationary time series random variables.
- Slow decay of ACF is usually an indication of potential non-stationarity (trend). Formal test of stationarity: Augmented Dickey-Fuller test (ADF Test) and Kwiatkowski-Phillips-Schmidt-Shin test (KPSS test).
- Any strong stationary time series **with existing means and autocovariances** is also weakly stationary. The other direction does not hold, in general. For example, if X_t is IID Cauchy, X_t is strictly stationary but not weakly stationary.

Extension: Stationarity and Gaussian Processes

- An exception, where weak and strong stationarity are equivalent are Gaussian processes.
- Definition: The sequence $\{X_t\}$ is said to be a **Gaussian process** if for every choice of times $t_1, \dots, t_k$ the joint distribution of $(X_{t_1}, \dots, X_{t_k})$ is multivariate normal.
- In particular, it is much stronger than saying that each of $X_{t_1}, \dots, X_{t_k}$ has a univariate normal distribution. Why?
- The multivariate normal distributions are uniquely determined by their means and covariances. Hence, we obtain

$$\text{Weak Stationarity} + \text{Gaussian Process} \implies \text{Strong Stationarity}$$

- We won't dive into Gaussian processes in this course any further though.

Section 4

Trend

Decomposing a time series into $signal(t)$ and noise

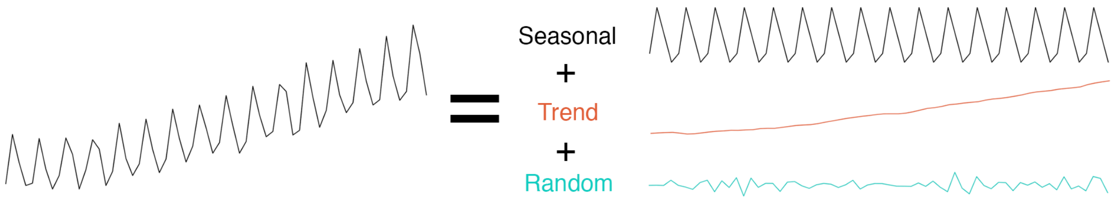


Figure 2: From anomaly.io

-
- We usually decompose the signal into a trend component "m_t" and a seasonal component "s_t":

$$signal(t) = m_t + s_t$$
$$\Rightarrow Y_t = m_t + s_t + X_t$$

This Unit: "Pursuing Stationarity"

- We will focus on the trend component this week and next week.
- We will discuss the seasonality afterwards.

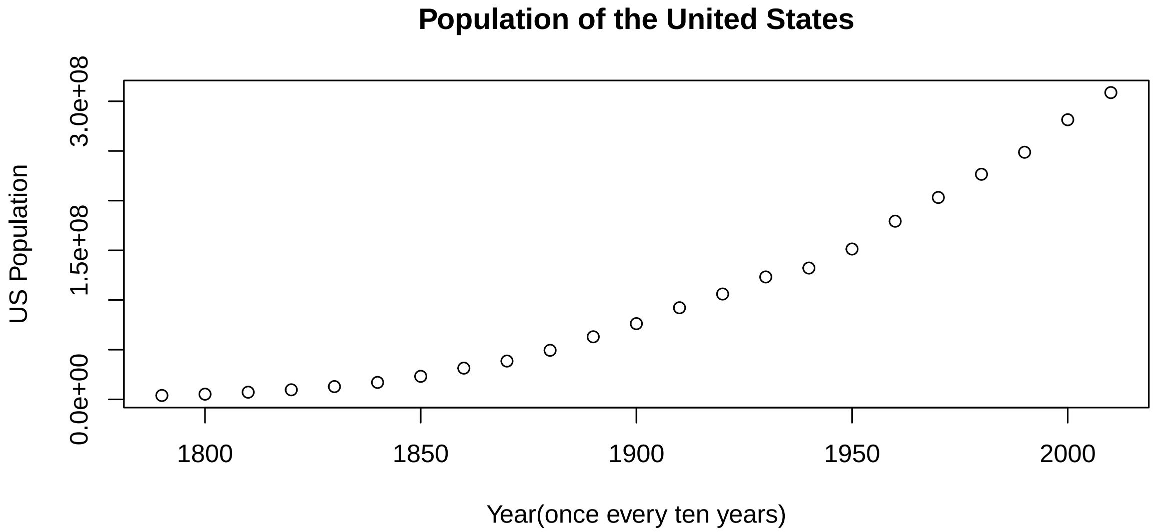
Trend Models

- As we'll look at seasonal effects later, our model today is

$$Y_t = m_t + X_t$$

- m_t is the trend
- X_t is as stationary process, perhaps white noise
- **Idea:** Model then remove the trend, so that data exhibits steady behavior over time, i.e. looks stationary. Then exploit dependence structure for estimation and prediction.

Example


Population of the United States
US Population
0.0e+00
1.5e+08
3.0e+08
1800
1850
1900
1950
2000
Year(once every ten years)

First Idea: Estimate trend $\hat{m}_t$.

If $\hat{m}_t \approx m_t$, then the residuals

$$y_t - \hat{m}_t \approx y_t - m_t = X_t$$

will have no trend over time.

Methods for estimating the trend:

- Parametric form for m_t, e.g., fit a polynomial with least squares
- Smoothing/Filtering - remove noise by averaging
- Other nonparametric methods - e.g. isotonic models, etc.

Section 5

Parametric Trend Estimation

Parametric Trend Estimation

- Assume parametric form for m_t, and estimate parameters.
- We will use additive linear models for this, where the variables can be any pre-defined functions of time:

$$m_t = \beta_0 + \beta_1 f_1(t) + \beta_2 f_2(t) + ... + \beta_p f_p(t)$$

- $f_j(t)$ can be any function: t, t^2, $log(t)$, $t * log(t)$, etc.

Least Squares

- Estimate the β parameters with least squares:

$$\hat{\beta} = \operatorname{arg\,min}_{\beta} \sum_t \left(Y_t - \beta_0 - \beta_1 f_1(t) - \dots - \beta_p f_p(t)\right)^2$$

- Lab this week will cover least squares, especially for those not familiar with it!

Example: Quadratic Curve/Parabola

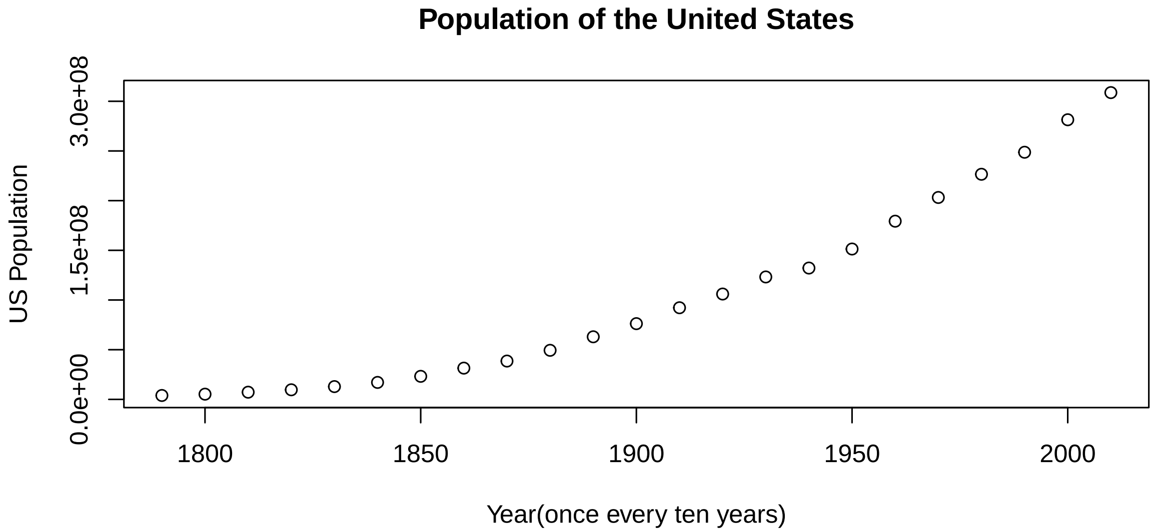
- Quadratic trend line: $m_t = \alpha + \beta t + \gamma t^2$
- Fit parameters with least squares

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \text{ argmin } \sum_t \left(Y_t - [\alpha + \beta t + \gamma t^2]\right)^2$$

then

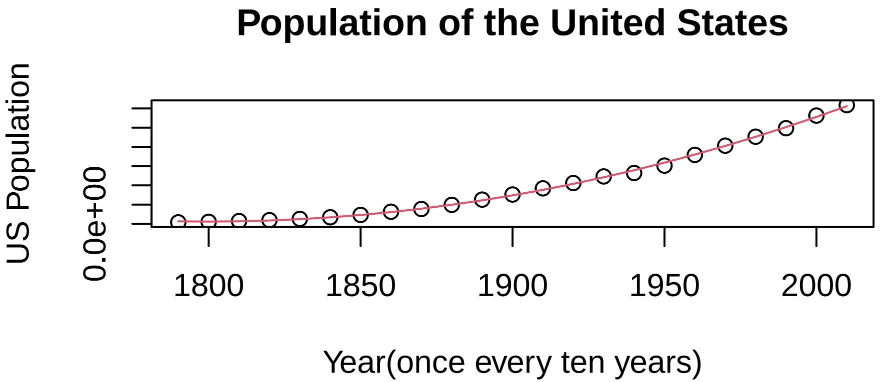
$$\hat{m}_t = \hat{\alpha} + \hat{\beta} t + \hat{\gamma} t^2$$

Example - US Population

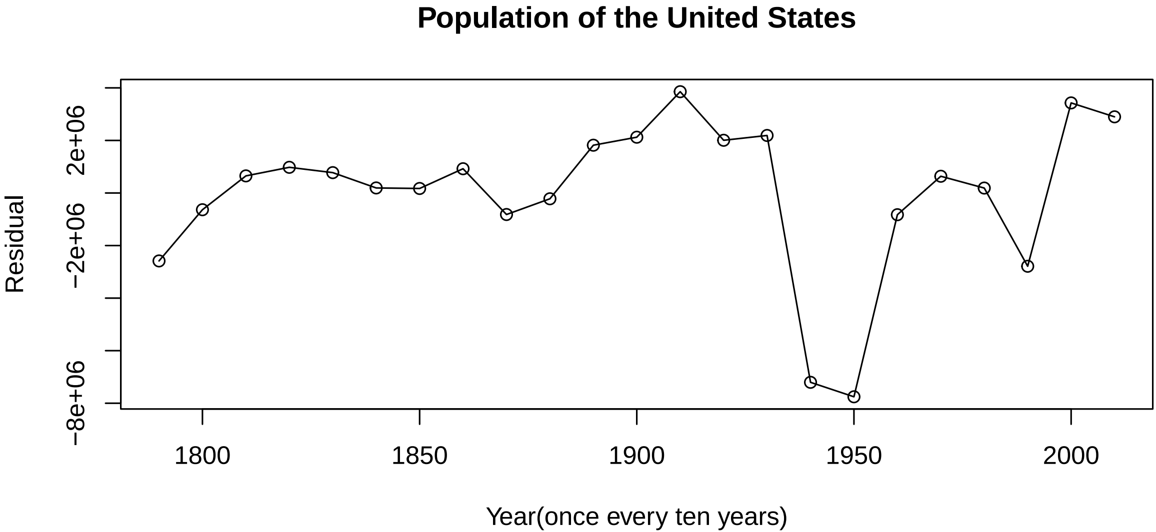


Example - with Trend

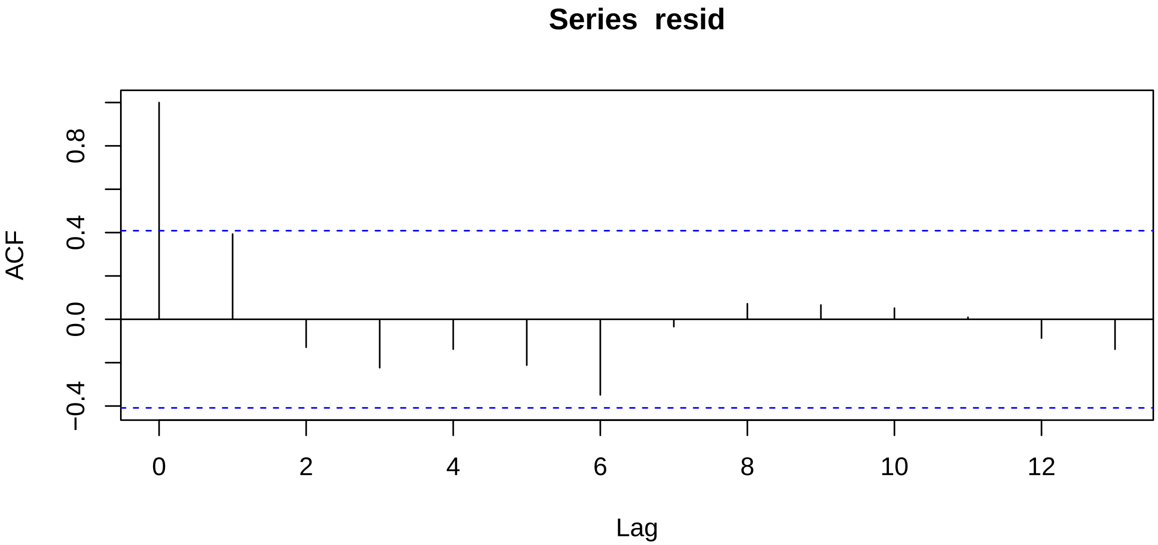
```
mod = lm(numpop ~ Year + I(Year^2),data=uspop)
plot(uspop$Year, uspop$numpop, type = "p", xlab = "Year(once every ten years)'
lines(uspop$Year,predict(mod),col=2)
```



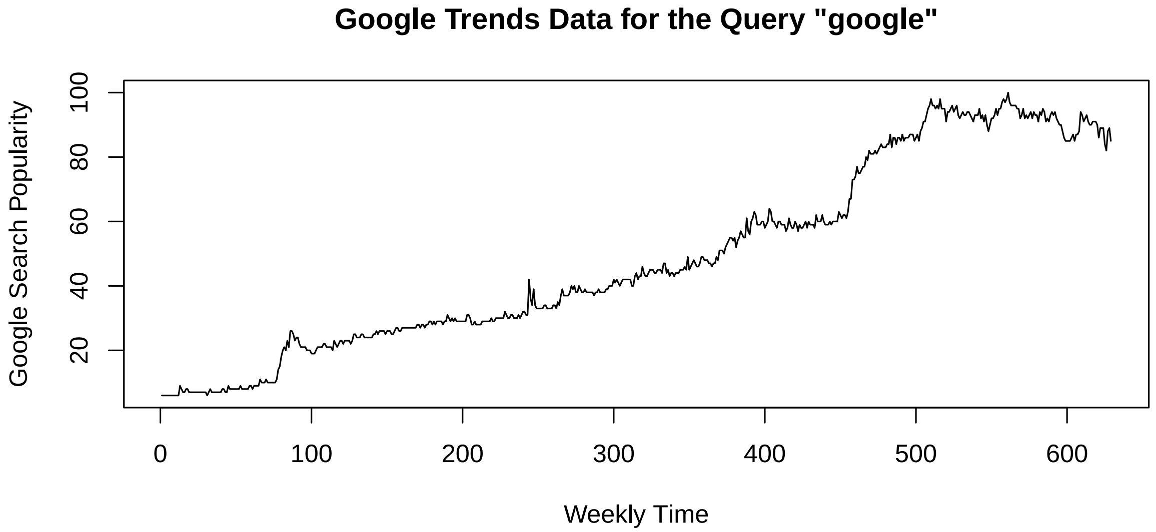
Example - Residuals



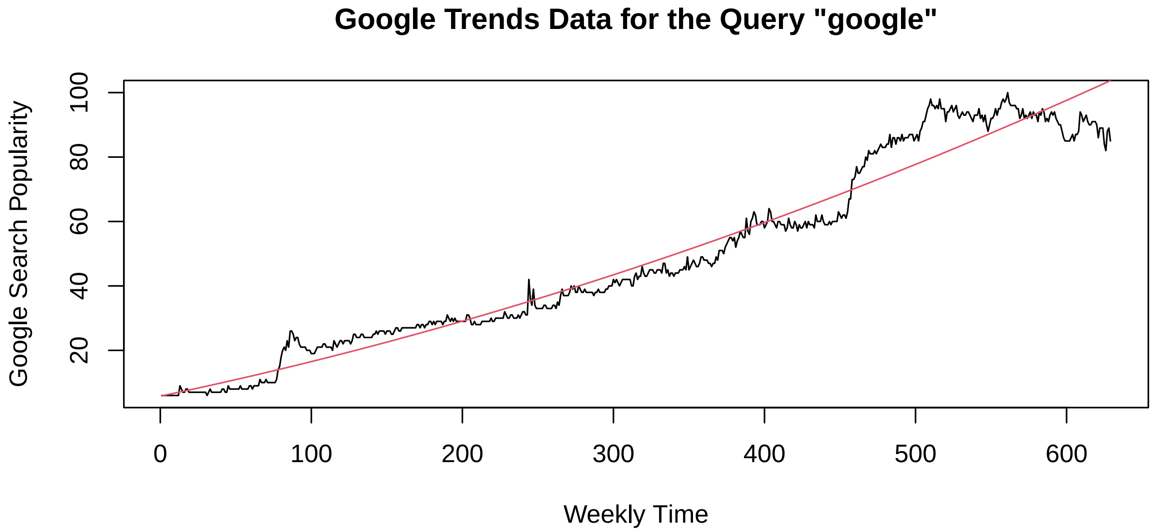
Example - ACF Correlogram of Residuals: $\hat{X}_t$ is plausibly white noise



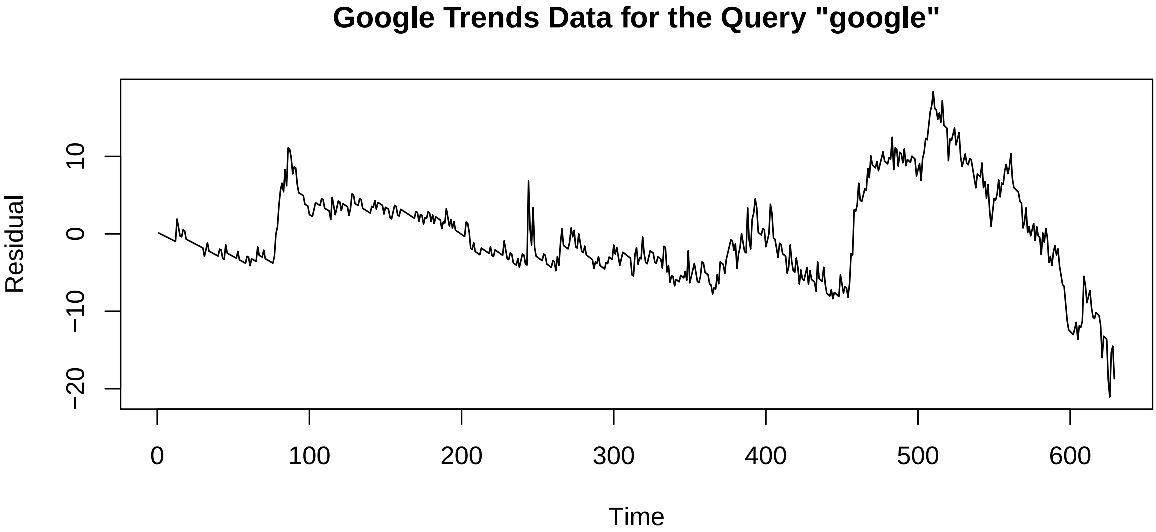
Example - Googling “Google”



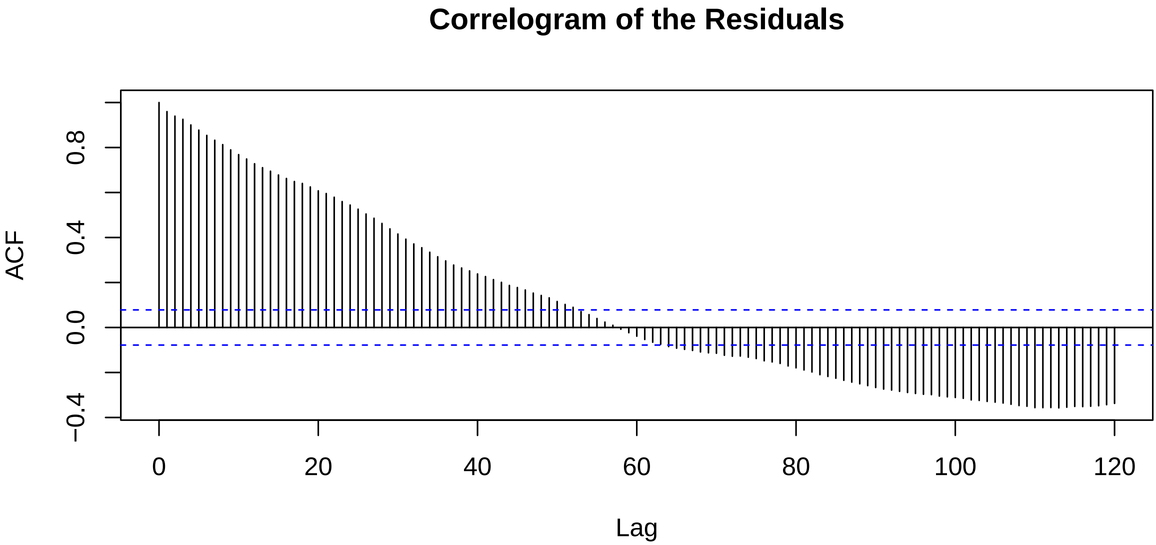
Example - with Trend



Example - Residuals



Example - ACF Correlogram of Residuals



Parametric Trend Estimation

Effectively deterministic in our model: better make sure it's plausible!

Advantages

- Gives very accurate estimates when model assumptions are correct.
- Straightforward to predict future observations.

Disadvantages:

- Selecting the correct model might be difficult.
- Parametric form might be unrealistic in practice.

Original Motivation:

Model and subtract the trend, so that the new series (the residuals) are steady over time (“reasonably stationary”)

Further Reading

https://anomaly.io/seasonal-trend-decomposition-in-r/index.html