Time Series 1

-Experimental Statistics II-

Lecturer: Darren Homrighausen, PhD

Goals

This exposition on time series is meant to give you a very brief and admittedly superficial overview

The goals of talking about time series in this class are:

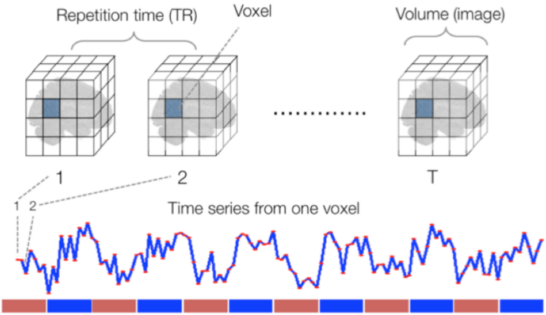
- to better understand the analysis of repeated measures
- to provide a foundation for exploring time series and other topics

  (These other topics will be imputation and Bayesian sampling)
- to try and allow you to answer yes in an interview setting to the question "do you have any experience with time series data"?

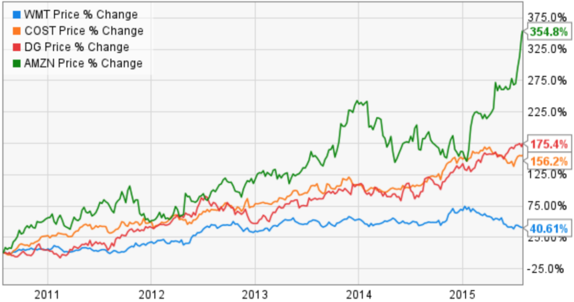
A reference for time series in SAS is
`http://www.statistik-mathematik.uni-wuerzburg.de/fileadmin/10040800/user_upload/time_series/the_book/2011-March-01-times.pdf`

Examples

Functional Magnetic Resonance Imaging (fMRI)



Stock market prices



Candy sales


Cadbury
creme egg


Cadbury DAIRY MILK Freddo
CLEARANCE
50% OFF
ALL EASTER STOCK

Overview

Key ingredients

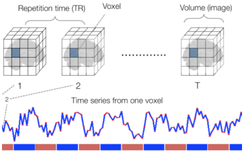
When considering time series data, it is customary to index with "t" instead of "i"

Let Y_t be the time series observed at time t

To start, let's presume that:

- Y_t is a univariate real number (that is, $Y_t \in \mathbb{R}$)
- there aren't any explanatory variables, which would be X_t

Example: In fMRI studies, the blood oxygenation level gets measured every 2 sec. Y_8 would be the 8^{th} measurement, made at 16 sec.



Key ingredients

There are two main parameters in this type of time series model:

Mean Function:

$$\mu_t = \mathbb{E} Y_t$$

Autocovariance Function:

$$\gamma(s,t) = cov(Y_s, Y_t) = \mathbb{E}[(Y_s - \mu_s)(Y_t - \mu_t)]$$

(Hence, the variance at time t is $\gamma(t,t)$)

Here, the added word "Function" is meant to convey that it is a function of t (time)

Example: Multiple regression

Let's connect the notation to multiple regression

Suppose $Y_t = \mu_t + \epsilon_t$
(Multiple regression would have $\mu_t = \beta_0 + X_t^\top \beta = \beta_0 + \sum_{j=1}^p X_{tj}\beta_j$)

Suppose ϵ_t are all independent $N(0, \sigma^2)$

Then:

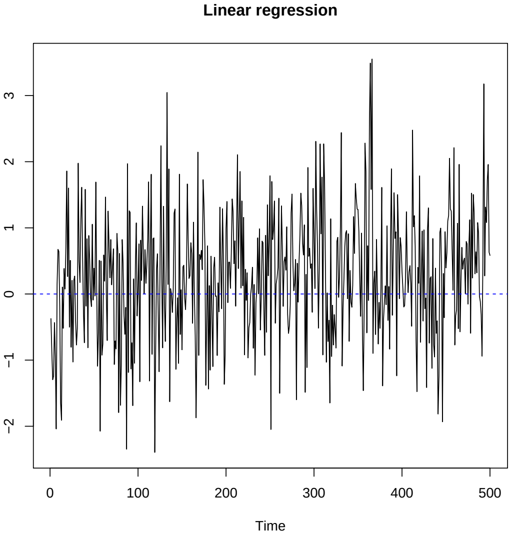
Mean Function:

$$\mathbb{E}Y_t = \mathbb{E}\mu_t + \mathbb{E}\epsilon_t = \mu_t$$

Autocovariance Function:

$$\gamma(s,t) = \mathbb{E}[(Y_s - \mu_s)(Y_t - \mu_t)] = \mathbb{E}\epsilon_s\epsilon_t = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$$

Example: Typical simulation



Example: Autoregression

Suppose $Y_t = 0.6Y_{t-1} + \epsilon_t$

Suppose ϵ_t are all independent $N(0, \sigma^2)$, and $Y_0 = 0$

Then:

Mean Function:

$$\mathbb{E}Y_t = \mathbb{E}[0.6Y_{t-1}] + \mathbb{E}\epsilon_t = 0$$

Autocovariance:

$$\gamma(s,t) = \mathbb{E}[(Y_s - \mu_s)(Y_t - \mu_t)] = \sigma^2 0.6^{|t-s|}$$

Covariance and Correlation

The terms covariance and correlation are almost the same thing

For two random variables U and V with $\mathbb{E}U = \mu_U$, $\mathbb{E}V = \mu_V$, $\mathbb{V}U = \sigma_U^2$, and $\mathbb{V}V = \sigma_V^2$:

- Covariance: For two random variables U and V their covariance is $\mathbb{E}[(U - \mu_U)(V - \mu_V)]$
- Correlation: For two random variables U and V their correlation is $\frac{\mathbb{E}[(U-\mu_U)(V-\mu_V)]}{\sigma_U \sigma_V}$

The only difference is that correlation has been normalized by the variance

Hence correlation must be in the interval $[-1, 1]$

Autocorrelation Function (ACF)

The autocovariance has the variance as a multiplier

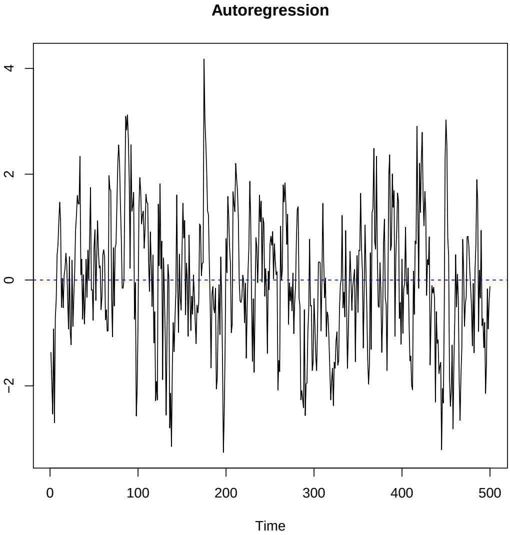
$$\gamma(s,t) = \mathbb{E}[(Y_s - \mu_s)(Y_t - \mu_t)] = \sigma^2 0.6^{|t-s|}$$

So, we can just divide by this to get a scaled version

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} = 0.6^{|t-s|}$$

Here, $\rho(s,t)$ is the autocorrelation function (ACF)

Example: Typical simulation



White noise

The noise terms ϵ_t are known as white noise in time series

(The etymology is that the electromagnetic spectrum of white light looks like the analogous representation for ϵ_t)

This is a very important component to time series, as it represents disturbances that are unforecasteable

(Often, the whole sequence $(\epsilon_t)_{t=1}^n$ is known as a white noise process)

Mean Function:

$$\mathbb{E}\epsilon_t = 0$$

Autocovariance Function:

$$\gamma(s,t) = \mathbb{E}[\epsilon_s \epsilon_t] \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$$

Autocorrelation Function:

$$\rho(s,t) = \begin{cases} 1 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$$

Regularity

A major topic in time series is stationarity

A time series $(Y_t)_{t=1}^n$ is stationary provided

- μ_t is constant over t
- The ACF depends on s, t only through $|t - s|$

Example:

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} = 0.6^{|t-s|}$$

As the ACF depends on s, t only through $|t - s|$, we might as well just write $t = s + h$

$$\rho(h) = \frac{\gamma(h)}{\sqrt{\gamma(0)\gamma(0)}} = 0.6^h$$

Estimation and Testing

Under stationarity, we can estimate μ and $\rho(h)$ for $h = 0, 1, 2, \ldots$
(Using sample averages)

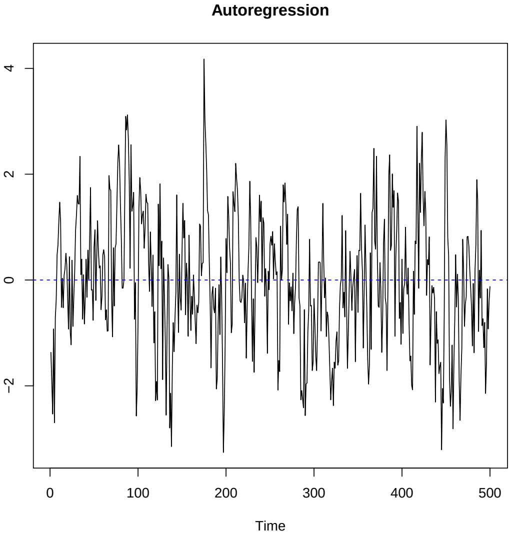
Write $\hat{\rho}(h)$ as the estimator of $\rho(h)$

Then $\hat{\rho}(h)$ has a particular distribution for white noise:

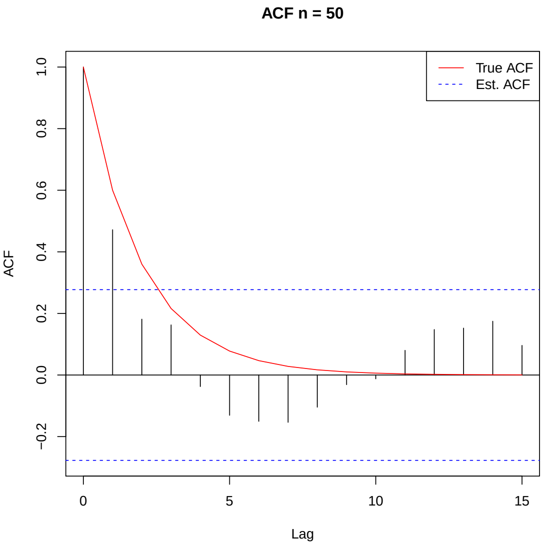
$$\hat{\rho}(h) \text{ is ``approx. distributed'' } \mathrm{Normal}(0, 1/\sqrt{n})$$

Hence, if we get a lot of values outside of, say, $\pm 2/\sqrt{n}$, a white noise process is unlikely

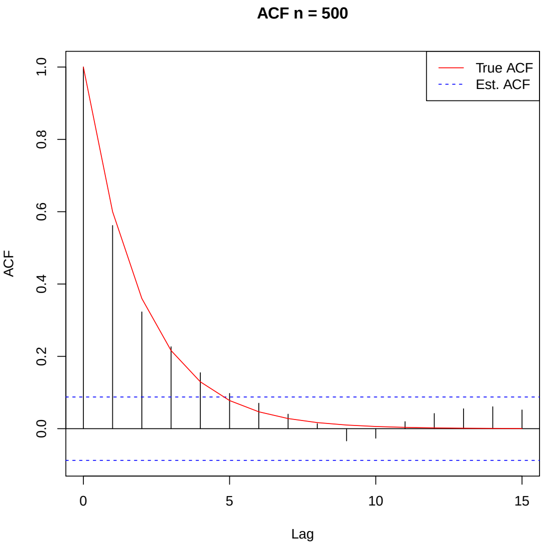
Data: autoregression



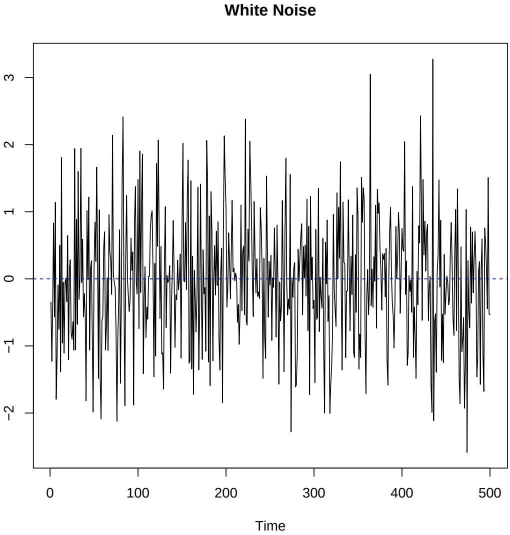
ACF: AUTOREGRESSION



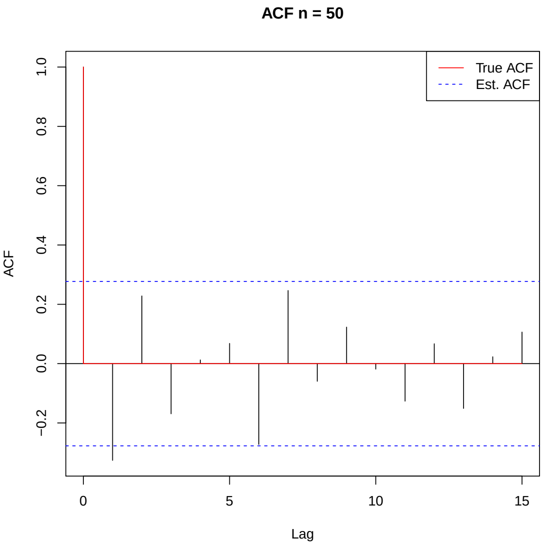
ACF: AUTOREGRESSION



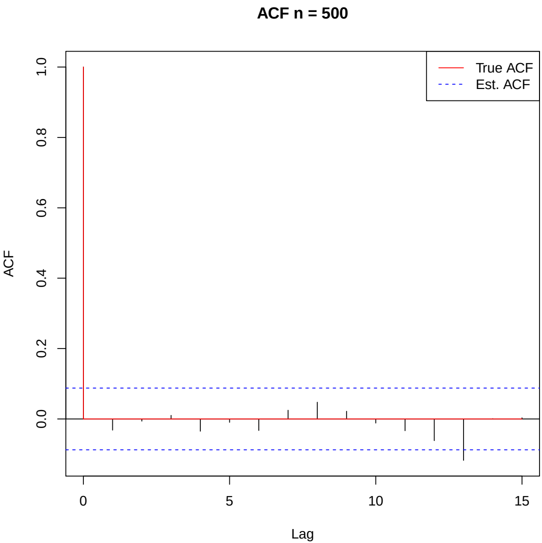
Data: white noise process



ACF: WHITE NOISE PROCESS



ACF: white noise process



ARIMA

Key processes

Autoregressive Process: An AR(p) process is:

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + \epsilon_t$$

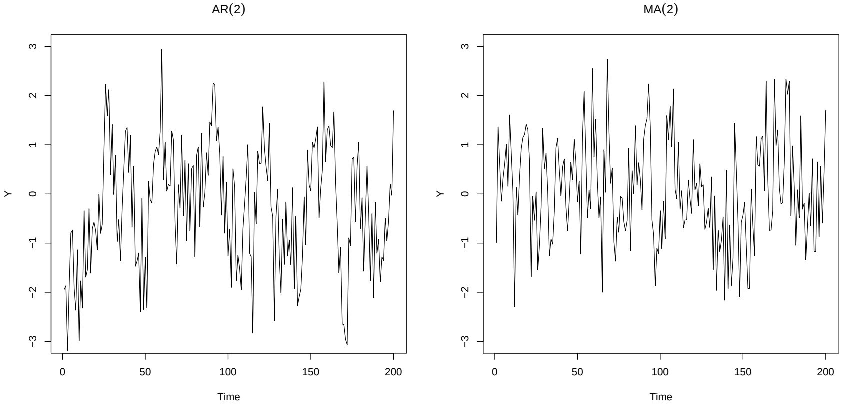
(This is a linear combination of the previous p values, plus a random noise term)

Moving Average Process: An MA(q) process is:

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} = \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

(This is a linear combination of the previous q random terms (sometimes referred to as shocks, plus a random noise term)

Examples: Some representative simulations



Key processes

ARMA process: A time series is an ARMA(p,q) process if it is stationary and a combination of AR and MA processes

It turns out the model is unidentified which makes it possible for a white noise process to look like a ARMA(k,k) process.

$$Y_t = \epsilon_t \Leftrightarrow Y_t - 0.5Y_{t-1} \underbrace{=}_{\text{using } Y_{t-1}=\epsilon_{t-1}} \epsilon_t - 0.5\epsilon_{t-1} \Leftrightarrow \text{ARMA}(1,1)$$

(Thus, we usually demand a stricter form of the ARMA model known as "causal" and "invertible")

Reminder: Autocovariance Function (ACF)

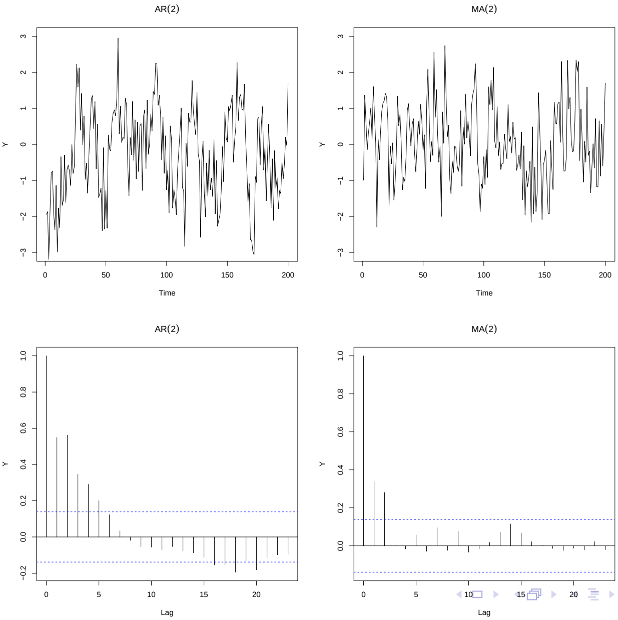
The ACF for a stationary process is:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

- AR(p): $\rho(h)$ will decay exponentially
- MA(q): $\rho(h)$ will have a sharp cut-off after $h > q$
- ARMA(p,q): $\rho(h)$ will decay exponentially

(The index h is usually known as a lag as it compares the value now to the value h units in the past/future)

Examples: Autocorrelation functions



ACF for ARMA models

Take away:

- The ACF gives a lot of information about the order of a MA process
- The ACF behavior is similar for AR and ARMA models
- The ACF gives little information about the order of a AR/ARMA process

To correct for these last two points, we need a different type of autocorrelation function

Partial Autocovariance (PACF)

For three random variables A, B, C, the partial covariance is

- Regressing A onto C to produce $\hat{A}$
- Regressing B onto C to produce $\hat{B}$
- Finding:

$$\text{partialcov}(A, B|C) = \text{cov}(A - \hat{A}, B - \hat{B})$$

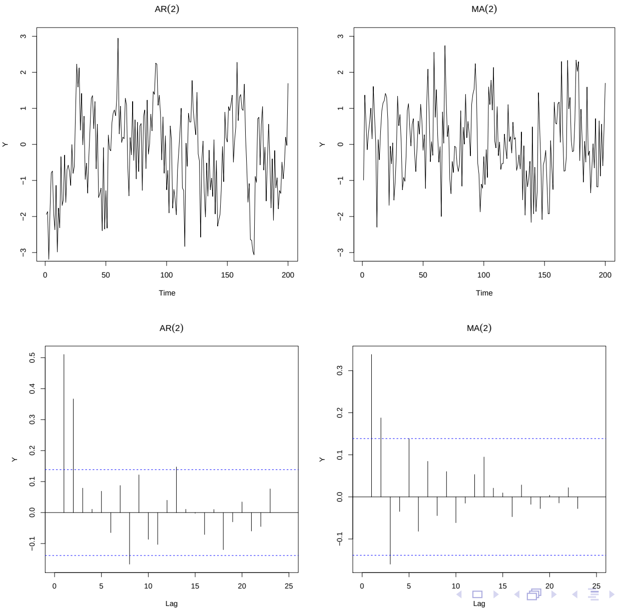
For an AR(p), Y_t is unrelated to Y_{t-p-1} given $\{Y_{t-1}, \ldots, Y_{t-p}\}$

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + \epsilon_t$$

$\rightarrow$ Use the partial correlation to determine the presence/order of AR processes

(The partial autocorrelation is the covariance between Y_t and Y_s with the linear effect of everything "in the middle" removed, divided by the variance)

PACF



Summary: Using ACF vs. PACF

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after q lags	Tails off
PACF	Cuts off after p lags	Tails off	Tails off

Lastly, there is something known as the "inverse autocorrelation function", which we will not discuss in this class

Differencing nonstationary time series

Integrated ARMA (ARIMA)

Example:

$$Y_t = Y_{t-1} + \epsilon_t$$

If we difference this series, we end up with a white noise process (and hence stationary)

$$Y_t - Y_{t-1} = \epsilon_t$$

It turns out, many time series are of this form:

- A nonstationary (trend or seasonal) component
- A zero-mean stationary component

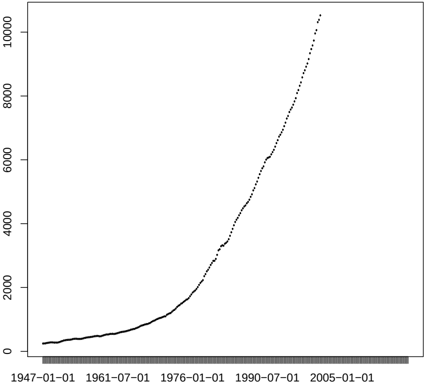
If differencing (Y_t) d-times produces an ARMA(p,q) process, then (Y_t) is an ARIMA(p,d,q)

An algorithm for ARIMA

1. Plot the data
2. Consider transformations
3. Identifying the dependence-based model parameters by looking at the PACF and ACF

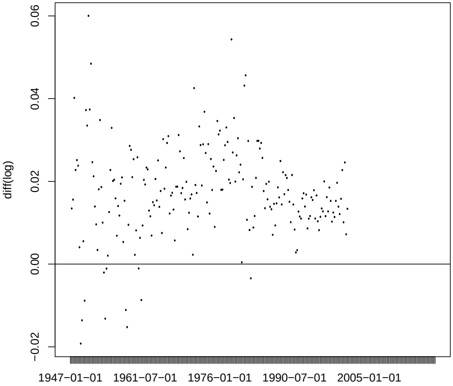
   (That is, p, d, and q)
4. Estimate the parameters
5. Check the residuals to see if they look like white noise
   - Look at the ACF/QQ plot of the residuals

United States GNP



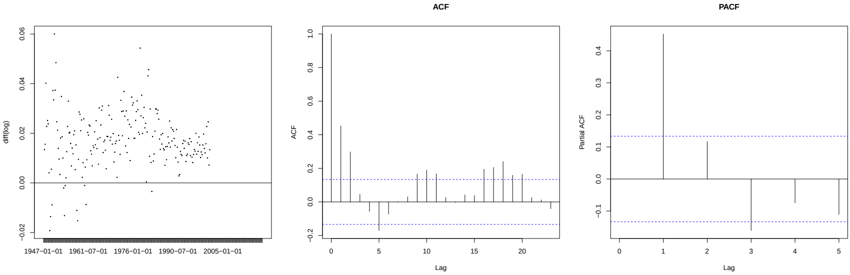
- The growth is probably exponential
- Definitely not stationary

United States GNP



- The growth is probably exponential $\rightarrow$ log transform
- Definitely not stationary $\rightarrow$ difference

United States GNP



- Could be an MA(2) process → we estimate that log(GNP) follows an ARIMA(0,1,2)
- Could be an AR(1) process → we estimate that log(GNP) follows an ARIMA(1,1,0)
- Could be a combination

We can fit all three models and compare..

Model selection

Though a topic in it's own right, we can select from competing models via Akaike's Information Criterion (AIC)

AIC can be viewed as estimating the distance from your estimated model and the true or best possible model

The way to use this criterion is to choose the model with the smallest value of AIC

(We would want to choose the model with the smallest distance, after all)