# Time-series analysis

Stefan Eng

5/12/2021

# Overview

1. Time series definitions
2. ARMA
3. Linear model approach
4. Alternative error models
5. Solutions

# Time series

- A time series is a set of observations $x_t$, each recorded at a specific time $t$.
- Can be discrete or continuous
- $x_1, x_2, x_3, \ldots, x_n$

# Example from ExOnc

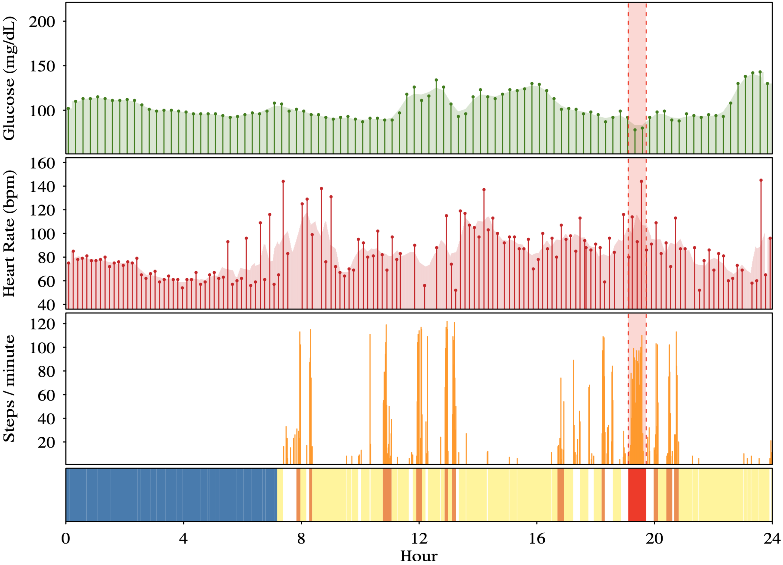



Figure 1: Visualization by Lydia Liu

# Time series definitions

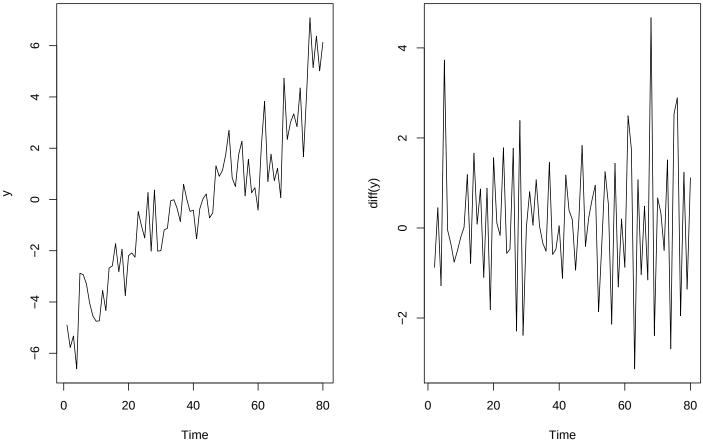
- Lag is a previous value in the time series
  - Lag 1 is previous value
  - Lag 2 for $x_t$ is $x_{t-2}$
- Autocorrelation/autocovariance function (`acf` function in R)
  - Computes the correlation for each lag in the time series
  - ACF at 0 is always 1
  - ACF at 1 is the correlation between $x_t$ and $x_{t-1}$ for all values in the time series
- Partial autocorrelation function (`pacf` function in R)

$$\frac{\text{Covariance}(x_t, x_{t-2} | x_{t-1})}{\sqrt{\text{Variance}(x_t | x_{t-1}) \text{Variance}(x_{t-2} | x_{t-1})}}$$

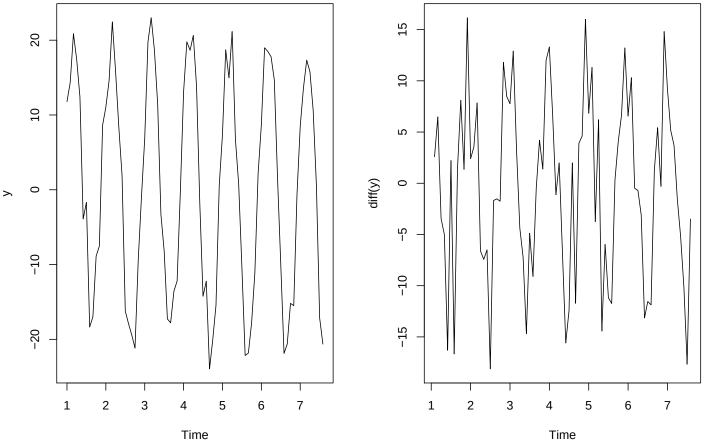
# Time series definitions

- **Differencing** a time series $\{x_t\}$ results in a new time series $\{x_t'\}$ where $x_t' = x_t - x_{t-1}$
- **Seasonality** is a repeating pattern that occurs on a period
- A time series is **stationary** if the ACF does not depend on time $t$, but only on the lag $h$.

# Differencing

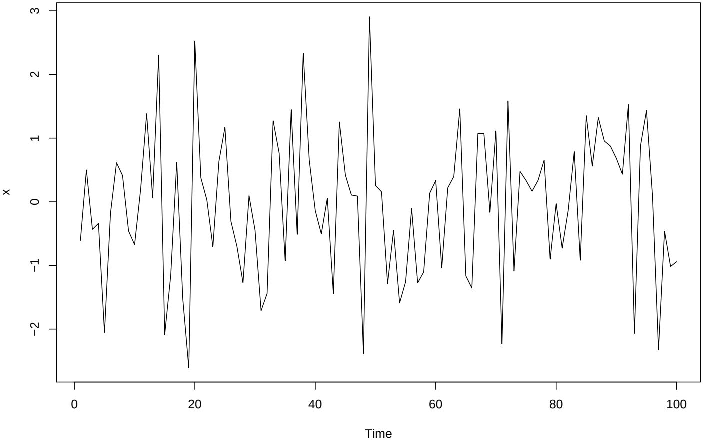


# Seasonality

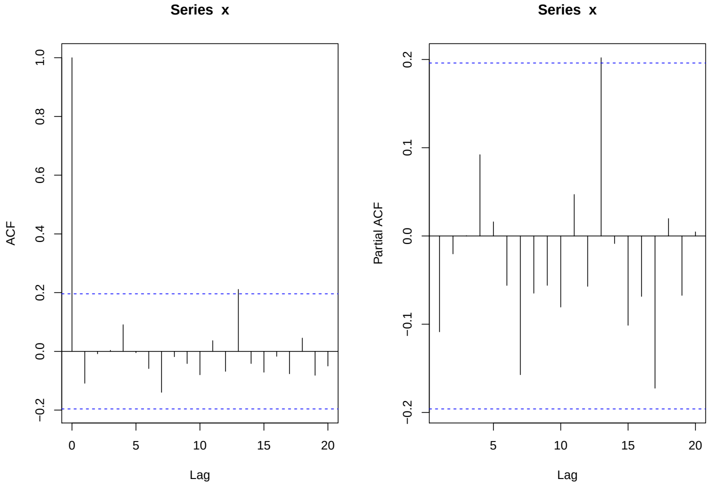


# Example - Normal

- $\{X_t\}, X_t \sim N(0, 1)$
- This is an iid sample so is no dependency between any of the points



# Example - Normal

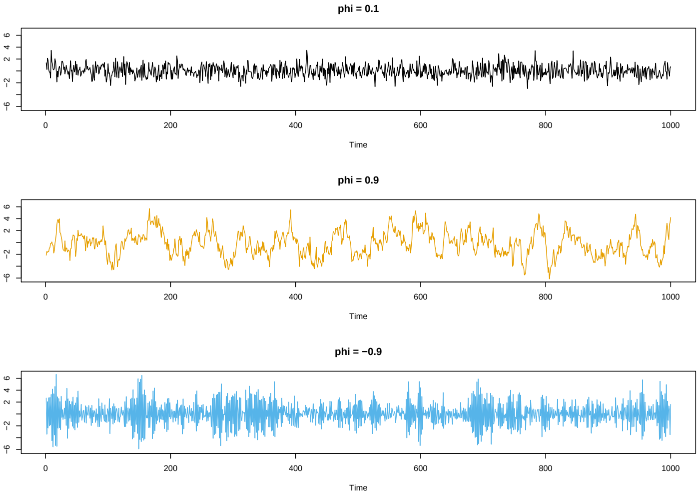


# Autoregressive model - AR(1)

- $X_t = \phi X_{t-1} + Z_t,\ t = 0, 1, \ldots$ where $Z_t$ is white noise, mean 0 sd $\sigma$, $|\phi| < 1$

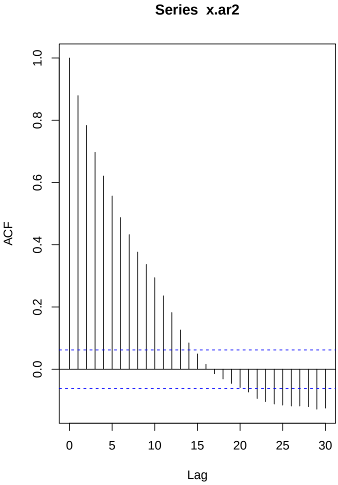
$$
\begin{aligned}
X_1 &= Z_1 \\
X_2 &= \phi Z_1 + Z_2 \\
X_3 &= \phi(\phi Z_1 + Z_2) + Z_3 \\
&= \phi^2 Z_1 + \phi Z_2 + Z_3 \\
X_t &= Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \cdots + \phi^{t-1} Z_1
\end{aligned}
$$

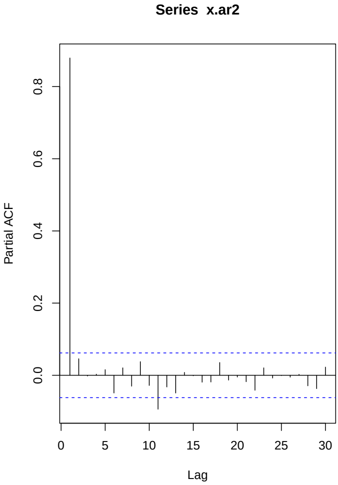
# Example - AR(1)


phi = 0.1
6
4
2
−2
−6
0
200
400
600
800
1000
Time
phi = 0.9
6
4
2
−2
−6
0
200
400
600
800
1000
Time
phi = −0.9
6
4
2
−2
−6
0
200
400
600
800
1000
Time

# Autoregressive AR(1) - ACF/PACF

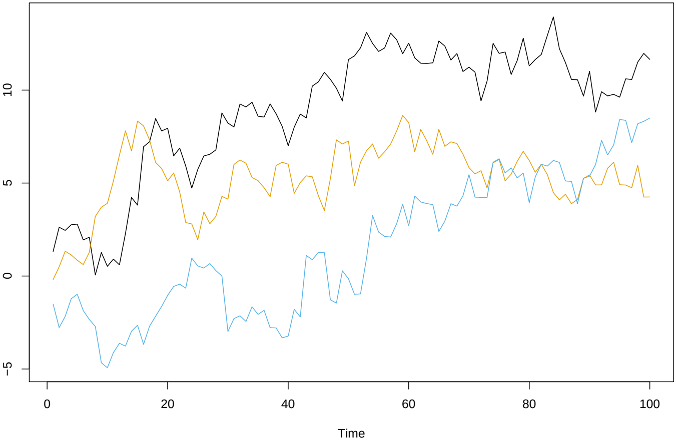
- In general, AR(p) should have **PACF** decreasing for $p$ lags then have 0 partial autocorrelation
- AR(p) should have **ACF** gradually decreasing to 0



# Examples - Random walk

If we let $\phi = 1$, then time series diverges and it is now a random walk



# Moving average - MA(1)

- MA(1): $X_t = Z_t + \theta Z_{t-1}$, $t = 0, 1, \ldots$ where $Z_t$ is white noise, mean 0 sd $\sigma$

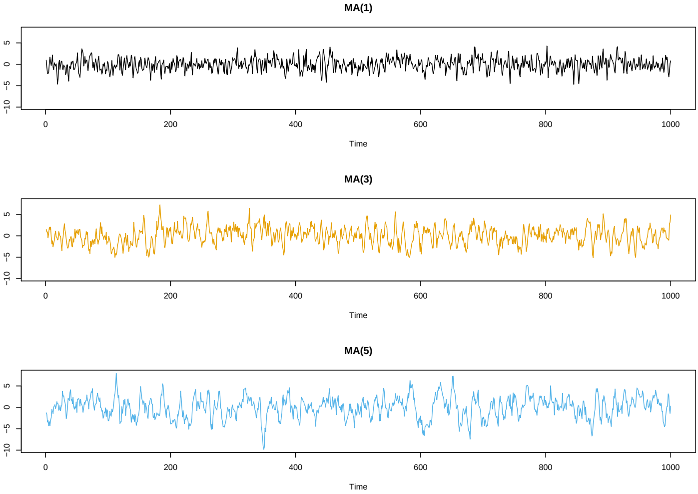
$$
\begin{aligned}
X_1 &= Z_1 \\
X_2 &= Z_2 + \theta Z_1 \\
X_3 &= Z_3 + \theta Z_2
\end{aligned}
$$

# Moving average - MA(3)

- MA(3): $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \theta_3 Z_{t-3}$, $t = 0, 1, \dots$
  where $Z_t$ is white noise, mean 0 sd $\sigma$

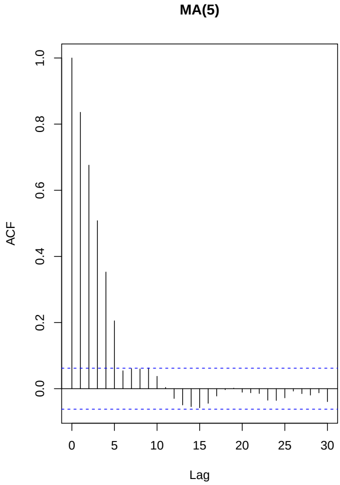
$$
\begin{aligned}
X_1 &= Z_1 \\
X_2 &= Z_2 + \theta_1 Z_1 \\
X_3 &= Z_3 + \theta_1 Z_2 + \theta_2 Z_1 \\
X_4 &= Z_4 + \theta_1 Z_3 + \theta_2 Z_2 + \theta_3 Z_1 \\
X_5 &= Z_5 + \theta_1 Z_4 + \theta_2 Z_3 + \theta_3 Z_2
\end{aligned}
$$

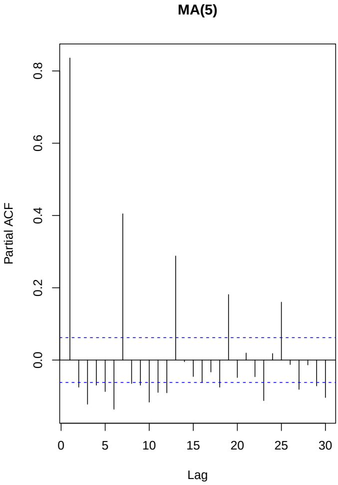
# Moving average MA(q)



# Moving average MA(q) - ACF/PACF

- In general, MA(q) should have ACF decreasing to $q$ lags then have 0 autocorrelation
- MA(q) should have PACF gradually decreasing to 0



# Simple Linear Regression

For a pair of $(x_i, y_i)$,

$$y_i = \beta x_i + \alpha + \epsilon_i$$

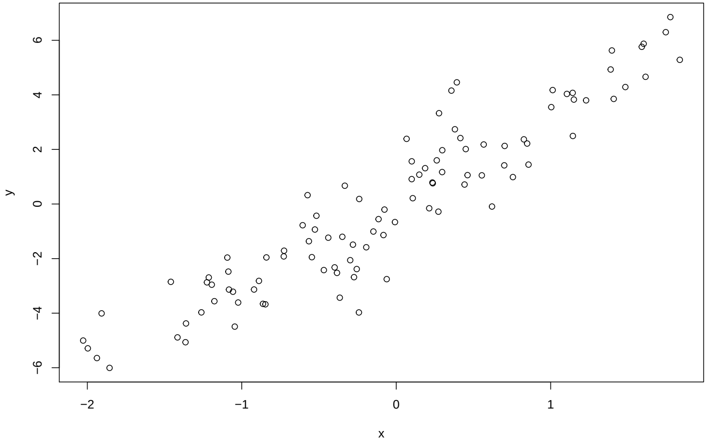
- $\varepsilon_i \sim N(0, \sigma^2)$
- $\{\varepsilon_i\}$ are **uncorrelated**
- variance is **constant**

# Simple Linear Regression - Time series

- If we have time series data $(t_i, y_i)$, the easiest way to model this would be to perform a linear regression of $y$ on $t$
  - Does this work with the assumptions? (Uncorrelated, constant variance)
- We can check the residuals (difference between predicted value and observed value)

# Simple Linear Regression - Time series

True model is $y_i = 3t_i + \varepsilon_i$, where $\{\varepsilon_i\}$ is AR(1).



# Simple Linear Regression - AR(1) Example

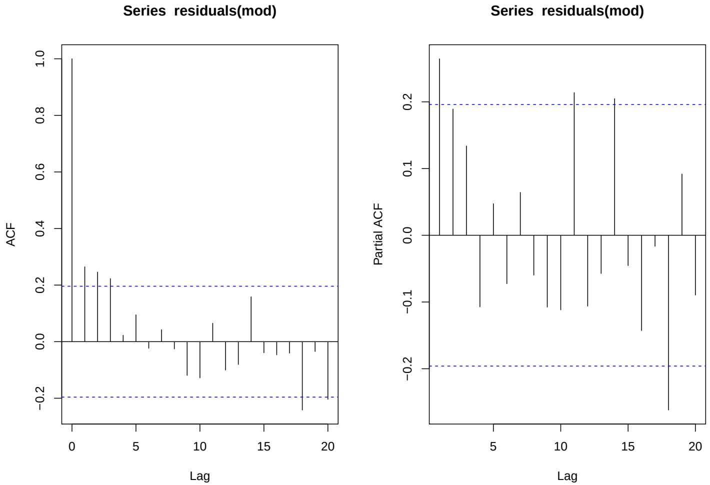
- First we can do the naive analysis with a linear model
- $\hat{\beta}$ (estimate for $\beta$) is still unbiased, even ignoring the error structure
- Standard errors are incorrect if we have misspecified the error structure

```
mod <- lm(y ~ x)
summary(mod)$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 0.1148917  0.1141530  1.006471 3.166676e-01
## x           3.0840179  0.1204029 25.614140 3.245706e-45
```

# Simple Linear Regression - AR(1) Example

```
par(mfrow = c(1, 2))
acf(residuals(mod))
pacf(residuals(mod))
```



# Simulations

- Simulate the following:
  - Linear regression when $y_i = 3t_i + \varepsilon_i$, where $\{\varepsilon_i\}$ is AR(1) with $\phi = 0.75$
  - Linear regression with standard normal errors
- $\text{Var}(\hat{\beta})$ is 13.68 times higher with the AR(1) errors than the variance in the correct model
- Note that the variance is not always higher, but could be biased in either direction
- In this case our standard errors will be higher on average and p-values higher as well

# Solutions

- Use a model that does not imply constant, non-correlated errors
  - `forecast::auto.arima` can fit linear regression with ARIMA errors
    - Coefficients can be added via `xreg` argument
  - `nlme::gls` generalized least squares fitting (Errors allowed to be correlated or unequal variance)
    - Correlation structure added via `correlation = corAR1(form = ~ 1 | group)`
  - `nlme::lme` linear mixed modeling (e.g. fixed and random effects)
    - Note that correlation structure modeling of the errors is not supported in `lme4`
- `mgcv` package: Generalized additive (mixed) models with support for correlation structure (Thanks Jaron for suggestion)

# auto.arima and gls

```r
nlme::gls(y ~ x, correlation = nlme::corAR1(form = ~ 1))
forecast::auto.arima(y, xreg = x, d = 0)
```

Questions?