EDS222 Week 9

Spatio-temporal Regression

November 25, 2024

Agenda

- **Autocorrelation**
  - What is it?
  - Why is it a problem and how do we diagnose it?
- **Solutions**
  - Depends on nature of autocorrelation
  - Lag models
  - Error models

Autocorrelation

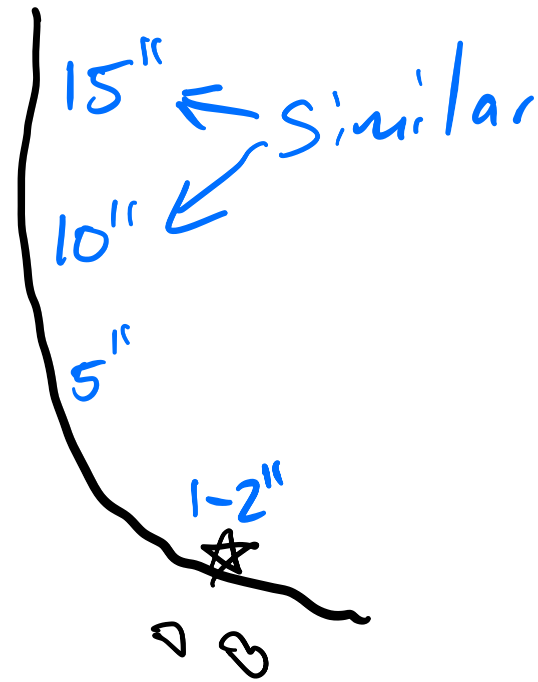
What is it?

Autocorrelation in time

Nearby events often similar

Autocorrelation

What is it?

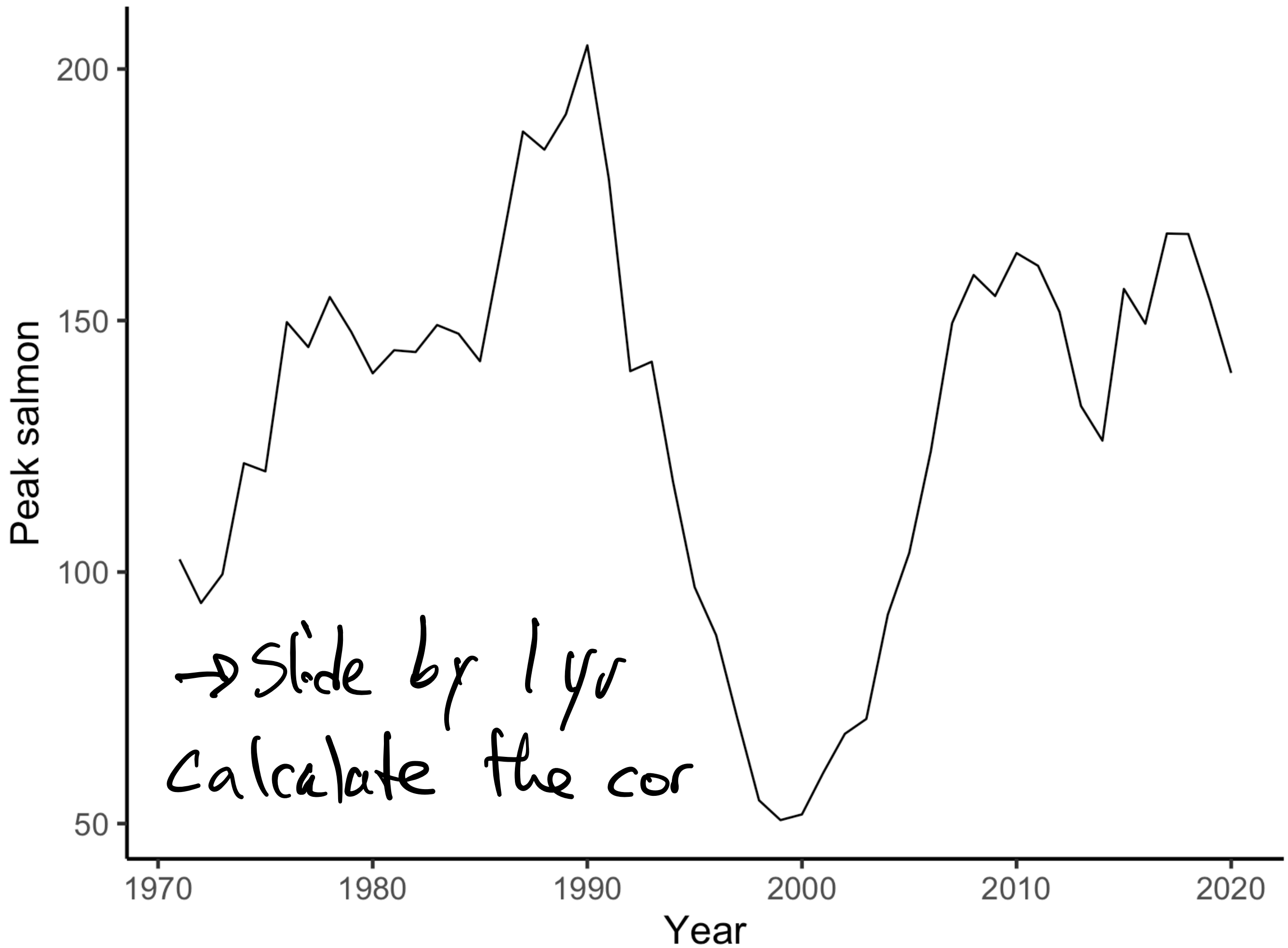


Spatial autocorrelation

Close in space = more similar

Autocorrelation

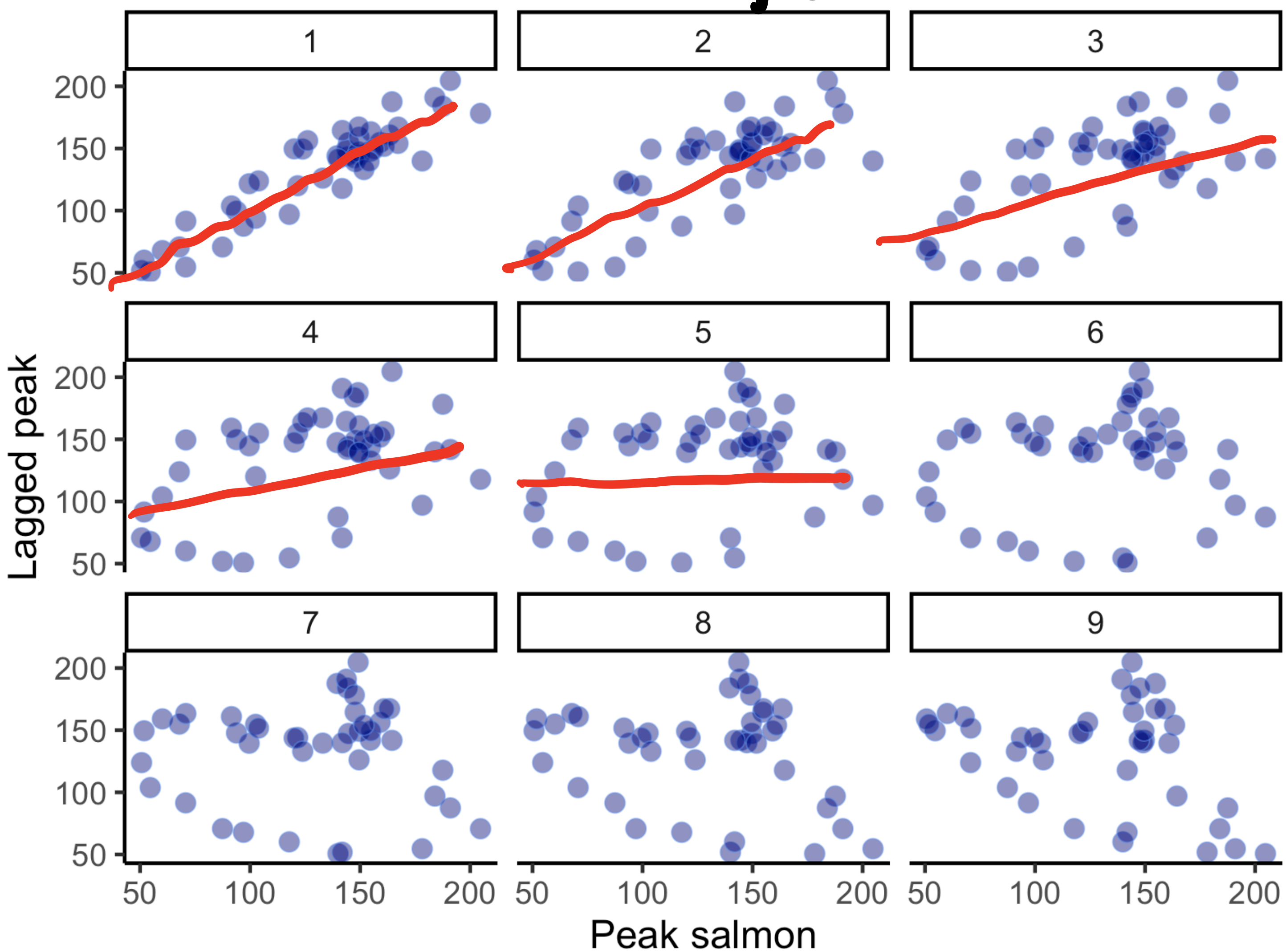
What is it?



Autocorrelation

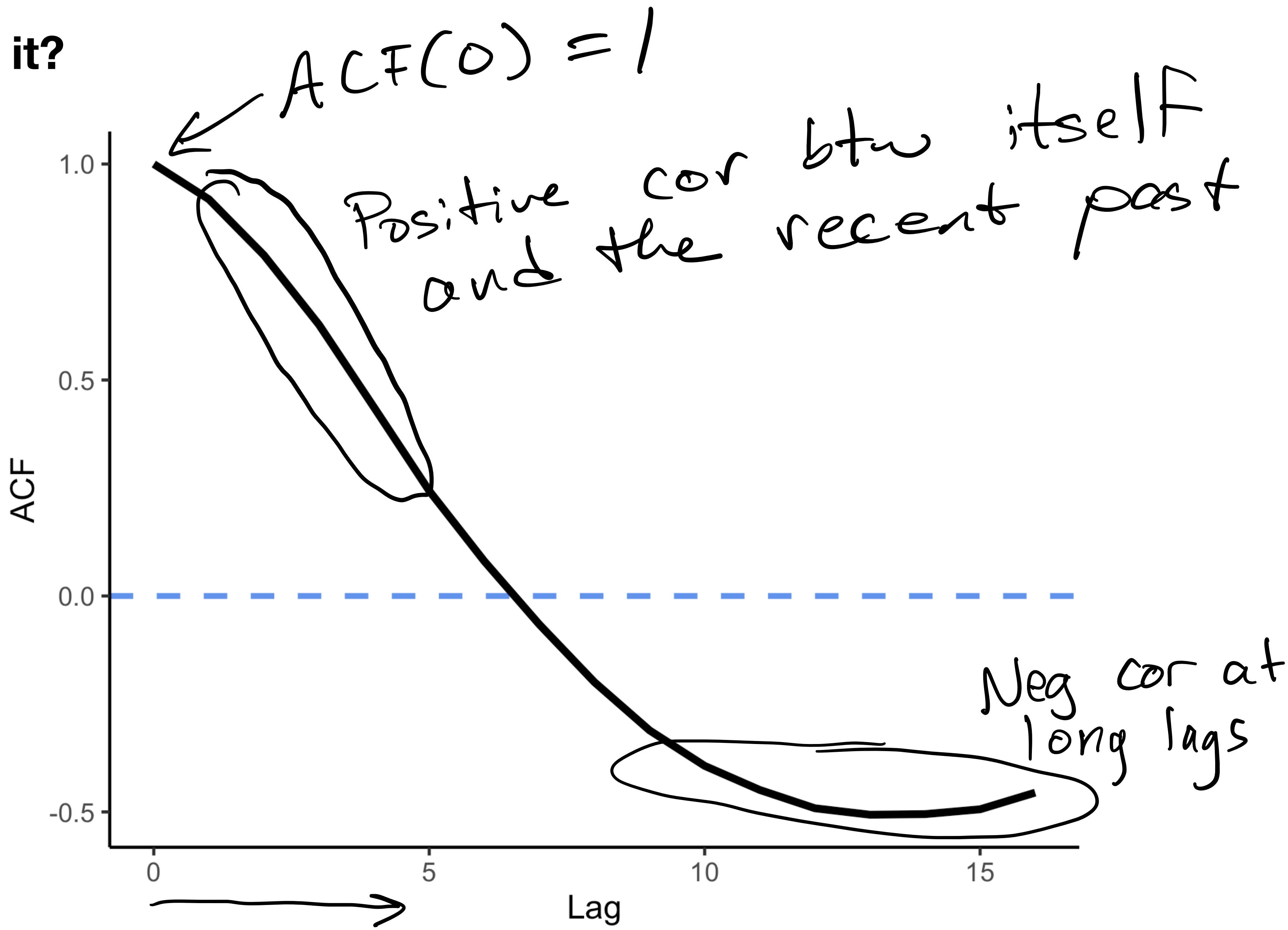
What is it?

Autocorrelation function ACF



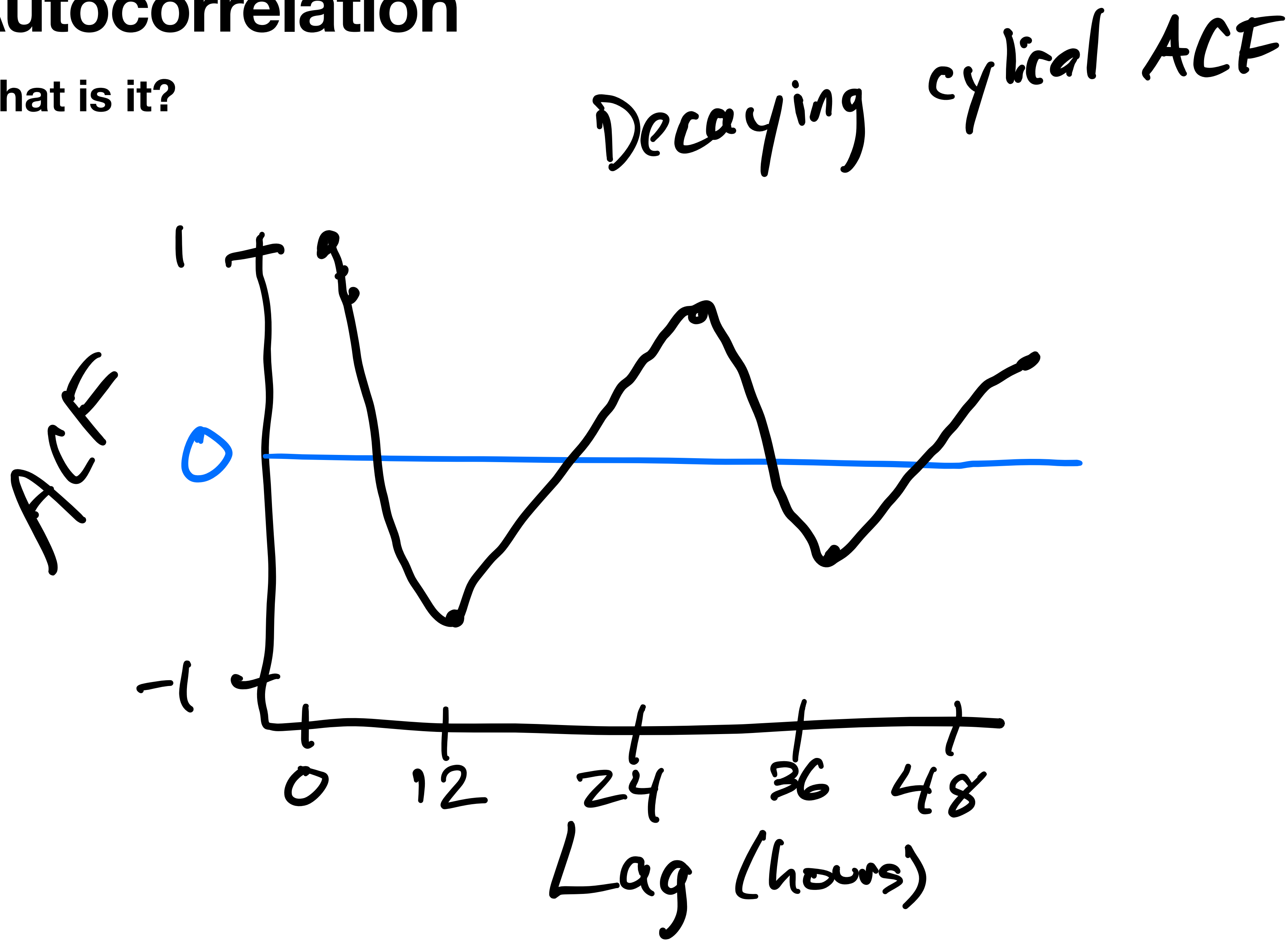
Autocorrelation

What is it?



Autocorrelation

What is it?



Autocorrelation

Why is it a problem and how do we diagnose it?

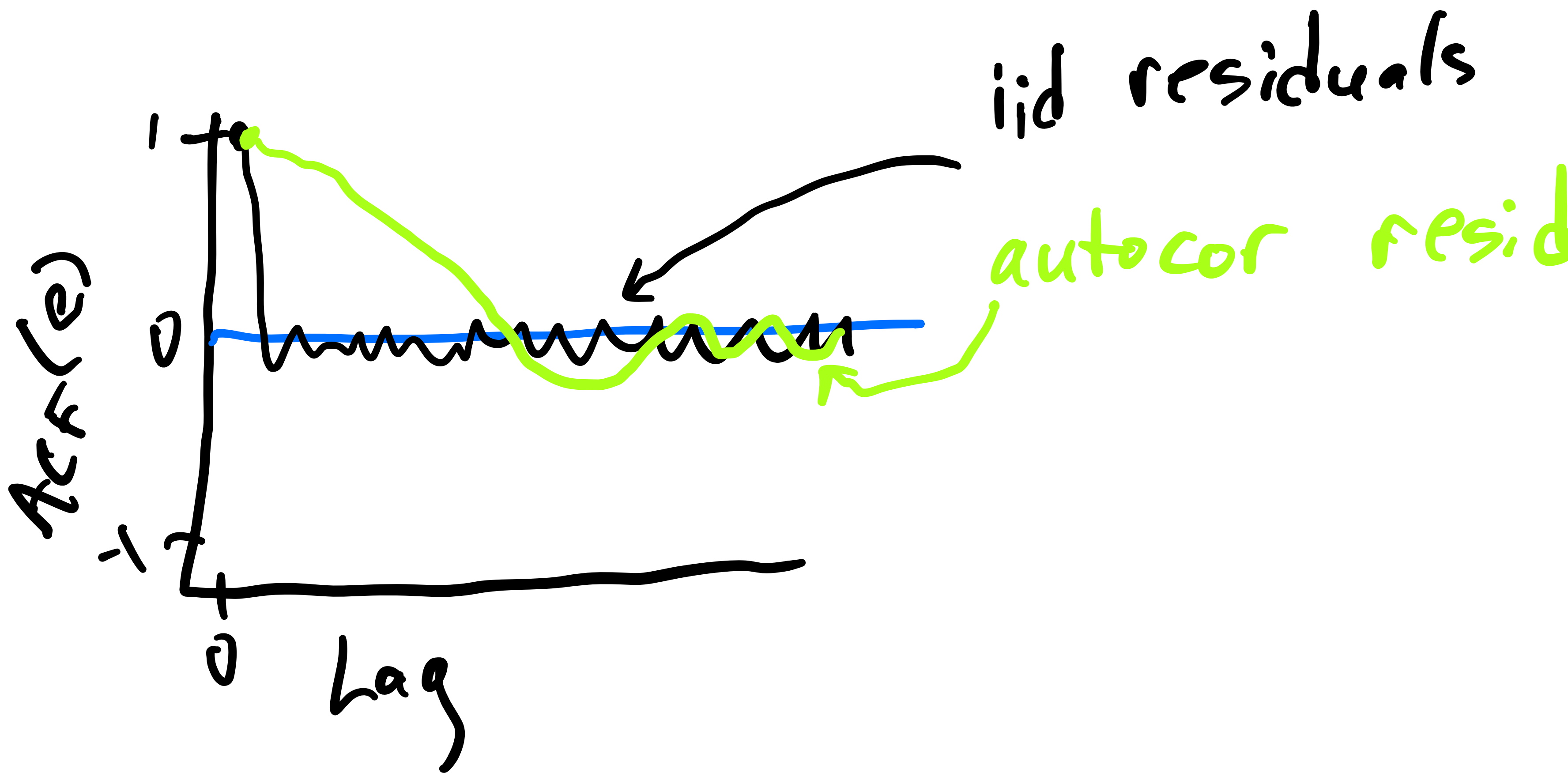
OLS assumption: "u is i.i.d. as normal"

Autocorrelation in your variables?
Maybe an issue

Autocorrelation in residuals?
BAD

Autocorrelation

Why is it a problem and how do we diagnose it?



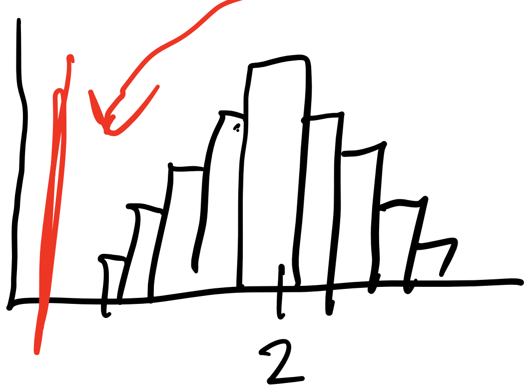
Autocorrelation

Why is it a problem and how do we diagnose it?

Hypothesis test!

1. H_0: No autocor in residuals
   H_A: Autocorrelated residuals

2. $D = \dfrac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$

3. 

4. Calculate your p-value

5. Interpret
   reject the null
   autocor resid

Autocorrelation

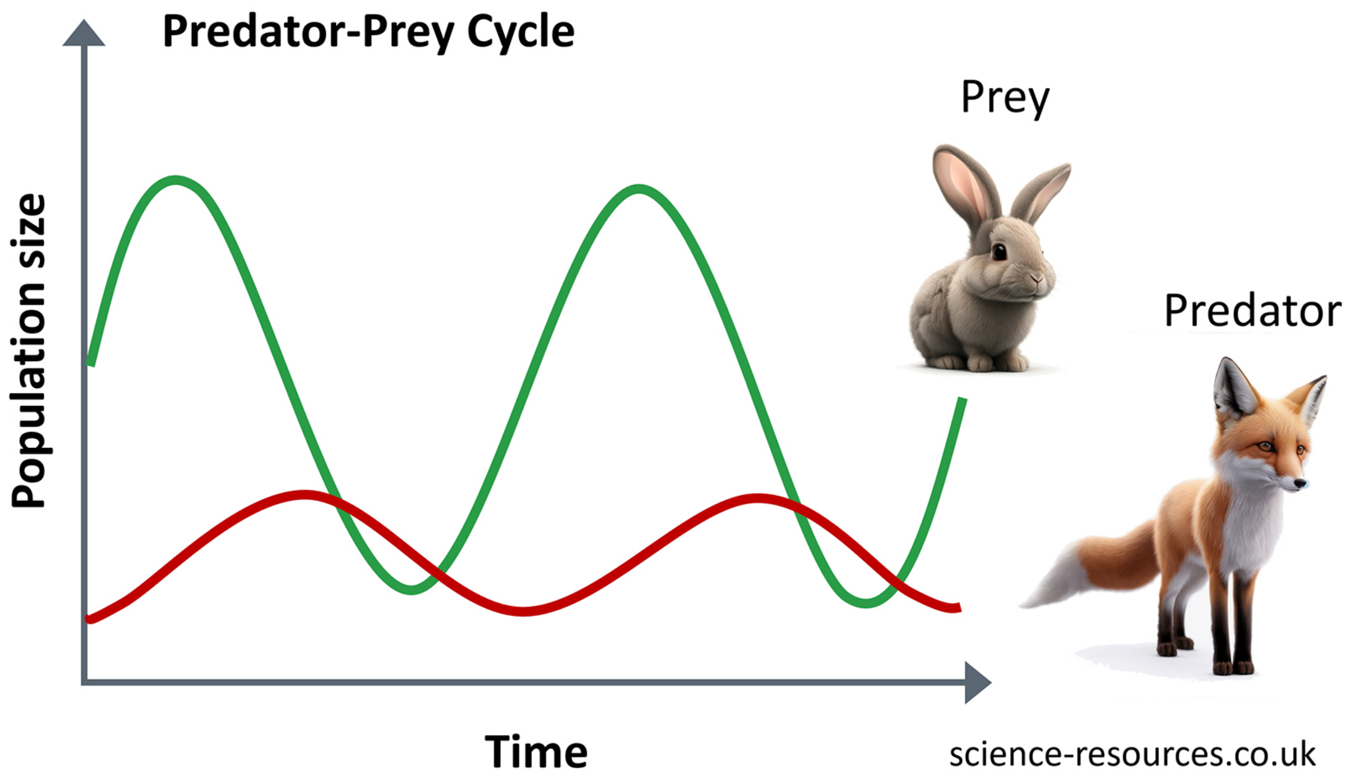
Recap

- Observations close in time or space are likely to be similar
- The autocorrelation function quantifies this self-similarity
- If residuals from a model are autocorrelated that means trouble

Solutions

Depends on nature of autocorrelation

1. Lags   Predictor and respone variables in the past influence the current value

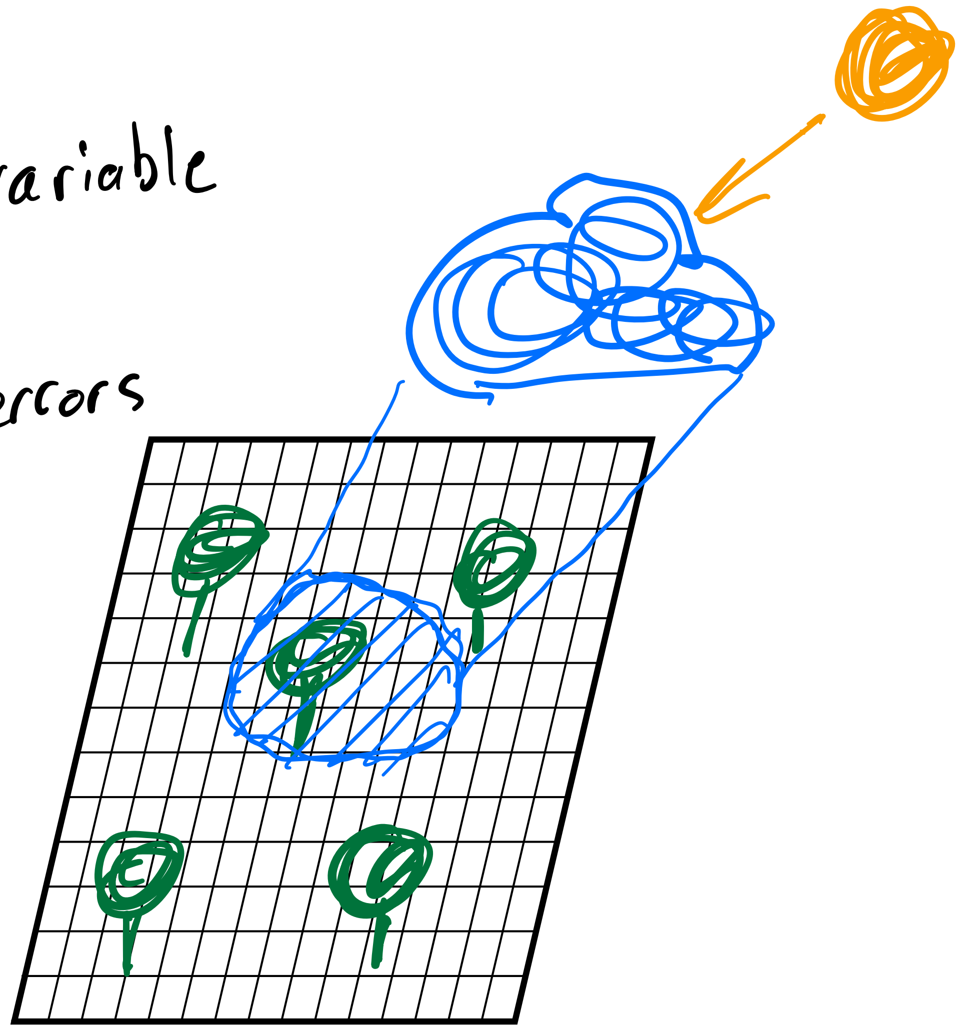


Solutions

Depends on nature of autocorrelation

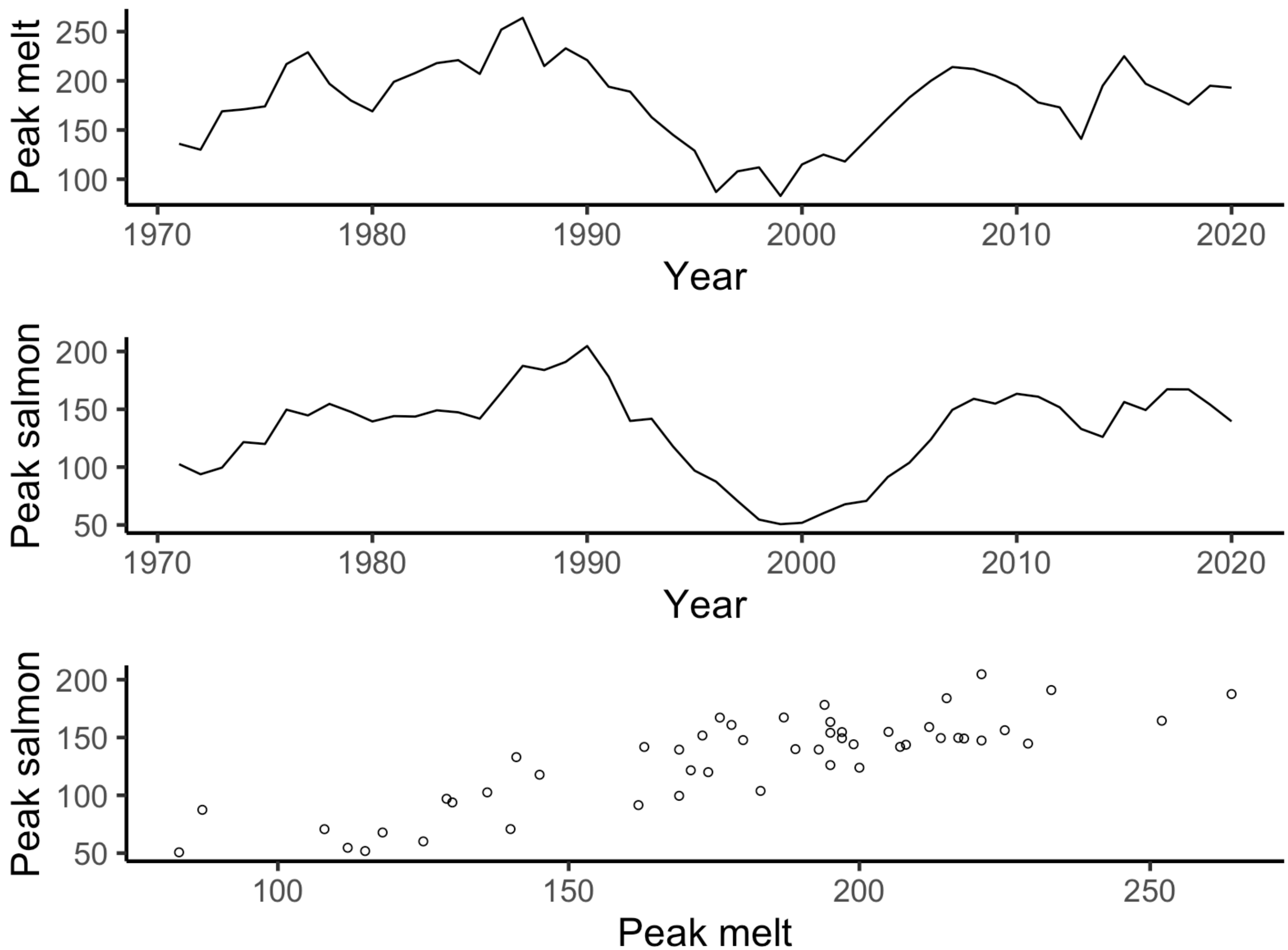
2. Errors

An unobserved variable is introducing autocorrelated errors directly



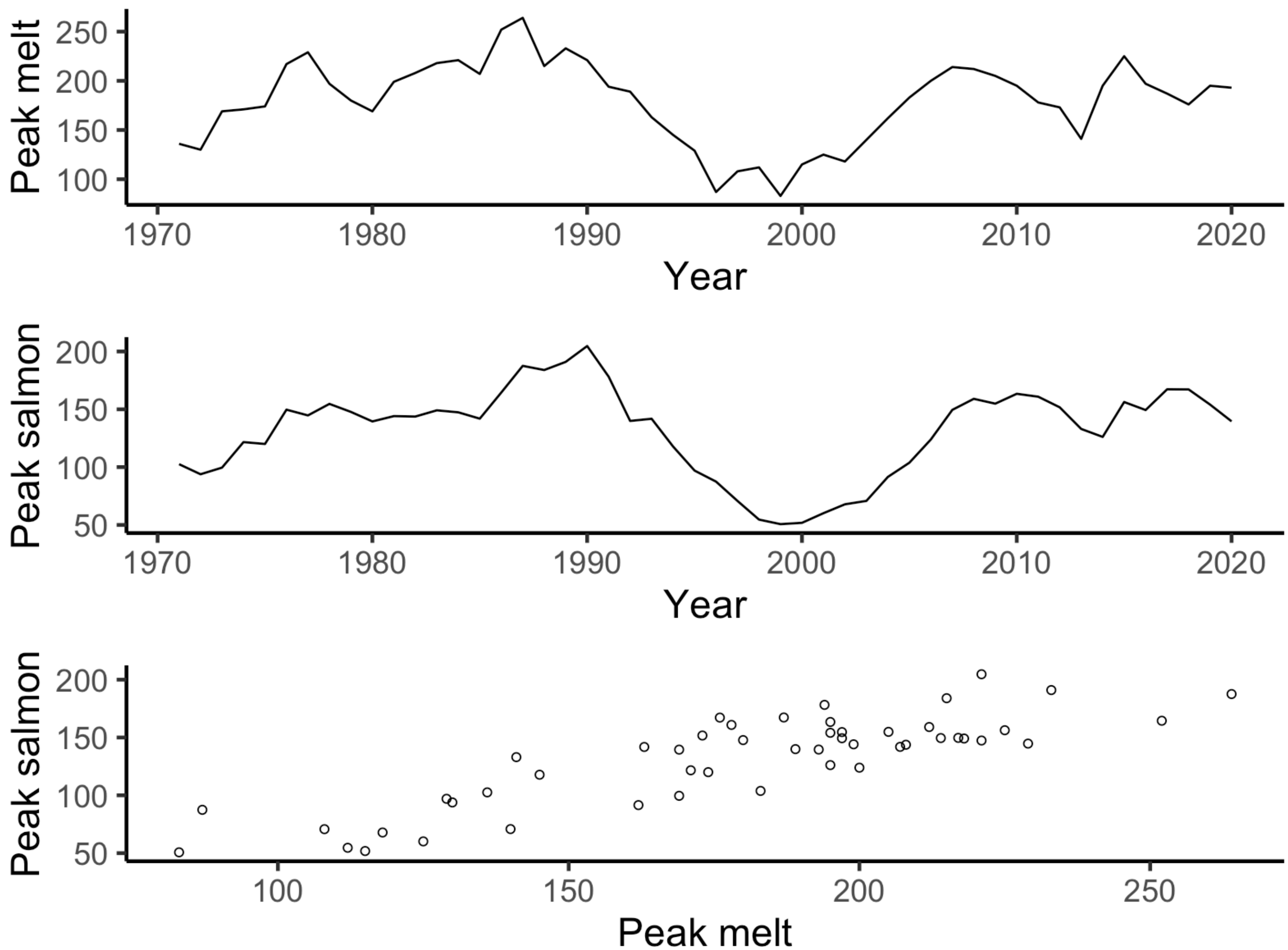
Solutions

Lag models



Solutions

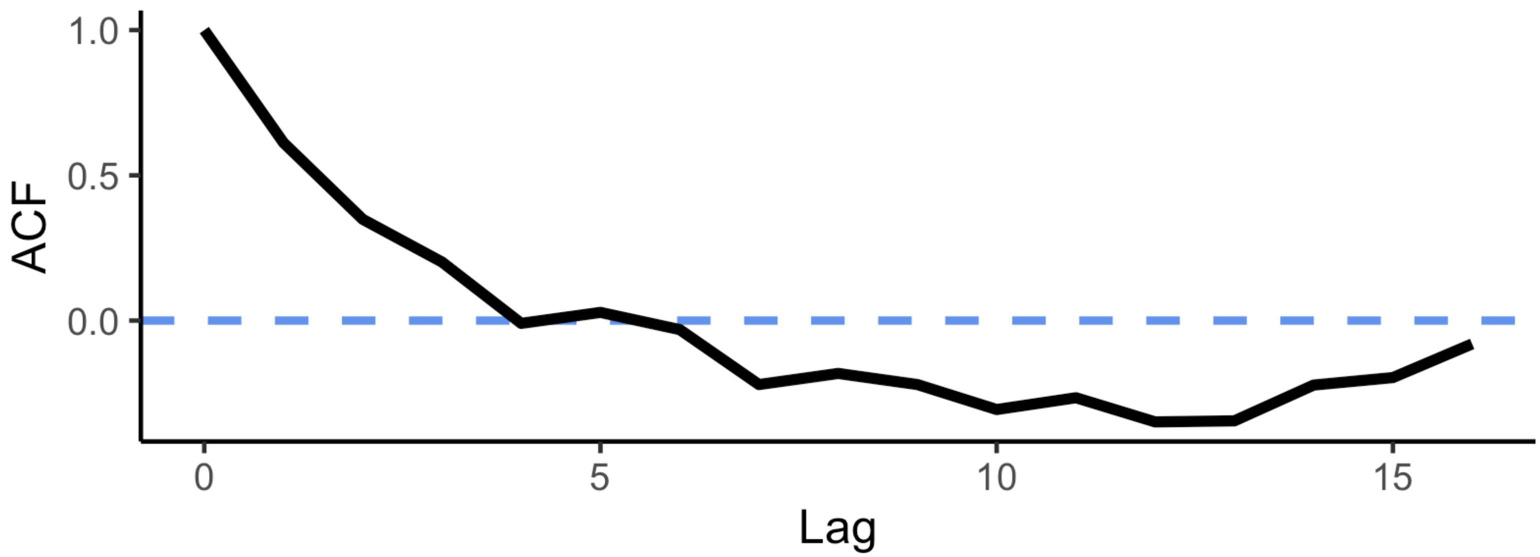
Lag models



Solutions

Lag models

```r
ols_mod <- lm(max_salmon_date ~ max_discharge_date, salmon_dat)
summary(ols_mod)
residual_acf <- acf(resid(ols_mod), plot = FALSE)
tibble(Lag = residual_acf$lag, ACF = as.vector(residual_acf$acf)) %>%
  ggplot(aes(Lag, ACF)) +
  geom_hline(yintercept = 0,
             linetype = "dashed",
             color = "cornflowerblue",
             linewidth = 1.5) +
  geom_line(linewidth = 2)
```



Solutions

Lag models

OLS: $Salmon = \beta_0 + \beta_1 \, melt + u$

Lag: $Salmon[t] = \beta_0 + \beta_1 \, melt[t] + \beta_2 \, Salmon[t-1] +$
$\beta_3 \, melt[t-1] + u$

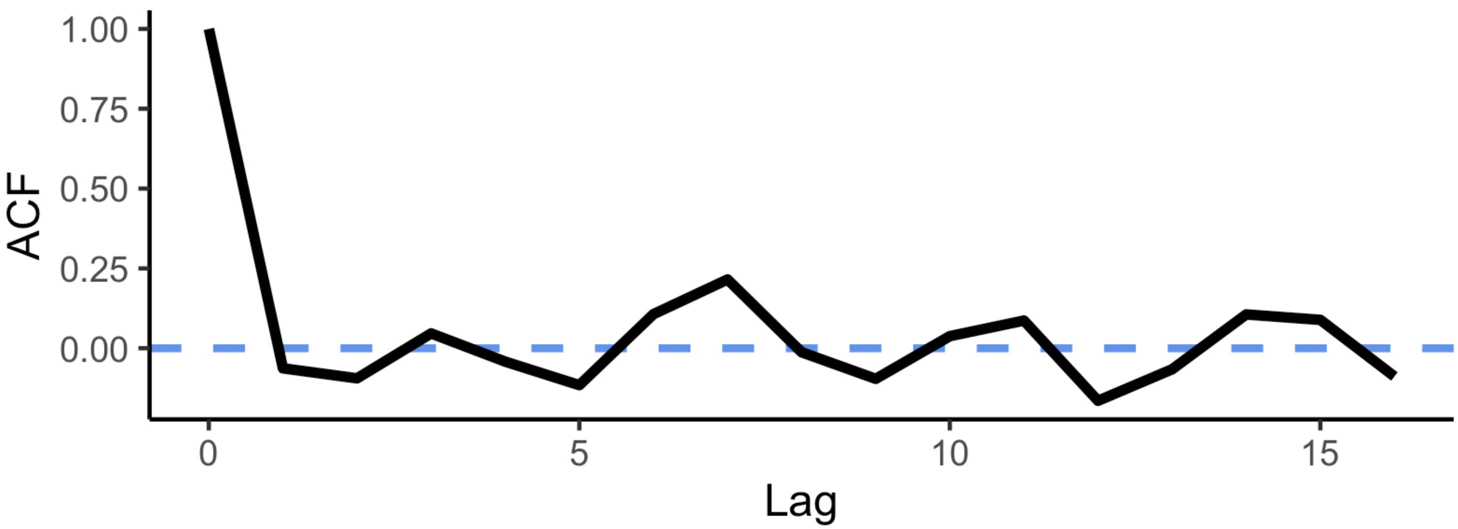
Autoregressive Distributed Lag Model

ADL(1, 1)

Solutions

Lag models

```
library(dynlm)
adl_mod <- dynlm(
  max_salmon_date ~ L(max_salmon_date, 1) + L(max_discharge_date, 0:1),
  ts(salmon_dat, start = 1971)
)
summary(adl_mod)
adl_resid_acf <- acf(resid(adl_mod), plot = FALSE)
tibble(Lag = adl_resid_acf$lag, ACF = as.vector(adl_resid_acf$acf)) %>%
  ggplot(aes(Lag, ACF)) +
  geom_hline(yintercept = 0,
             linetype = "dashed",
             color = "cornflowerblue",
             linewidth = 1.5) +
  geom_line(linewidth = 2)
```



Solutions

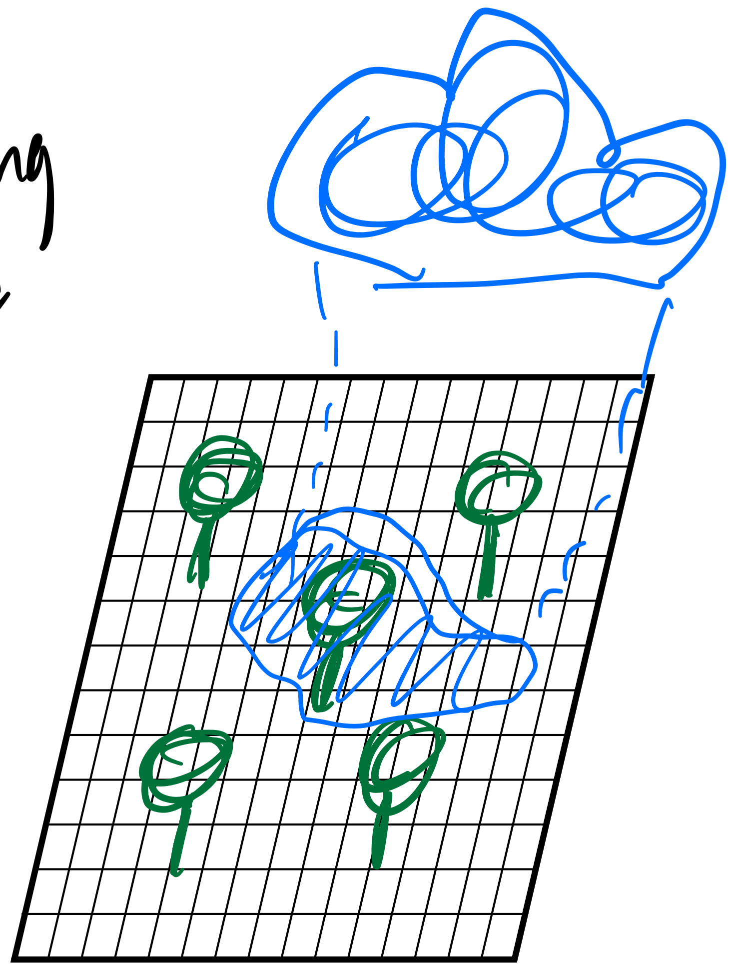
Lag models

- Lag models *lag* the response and/or predictor variables
- Works if the variables are the source of the autocorrelation
- Violates OLS assumptions - standard errors need to be handled with care
- Doesn’t work if there’s a trend
- Also applicable to spatial models!

Solutions

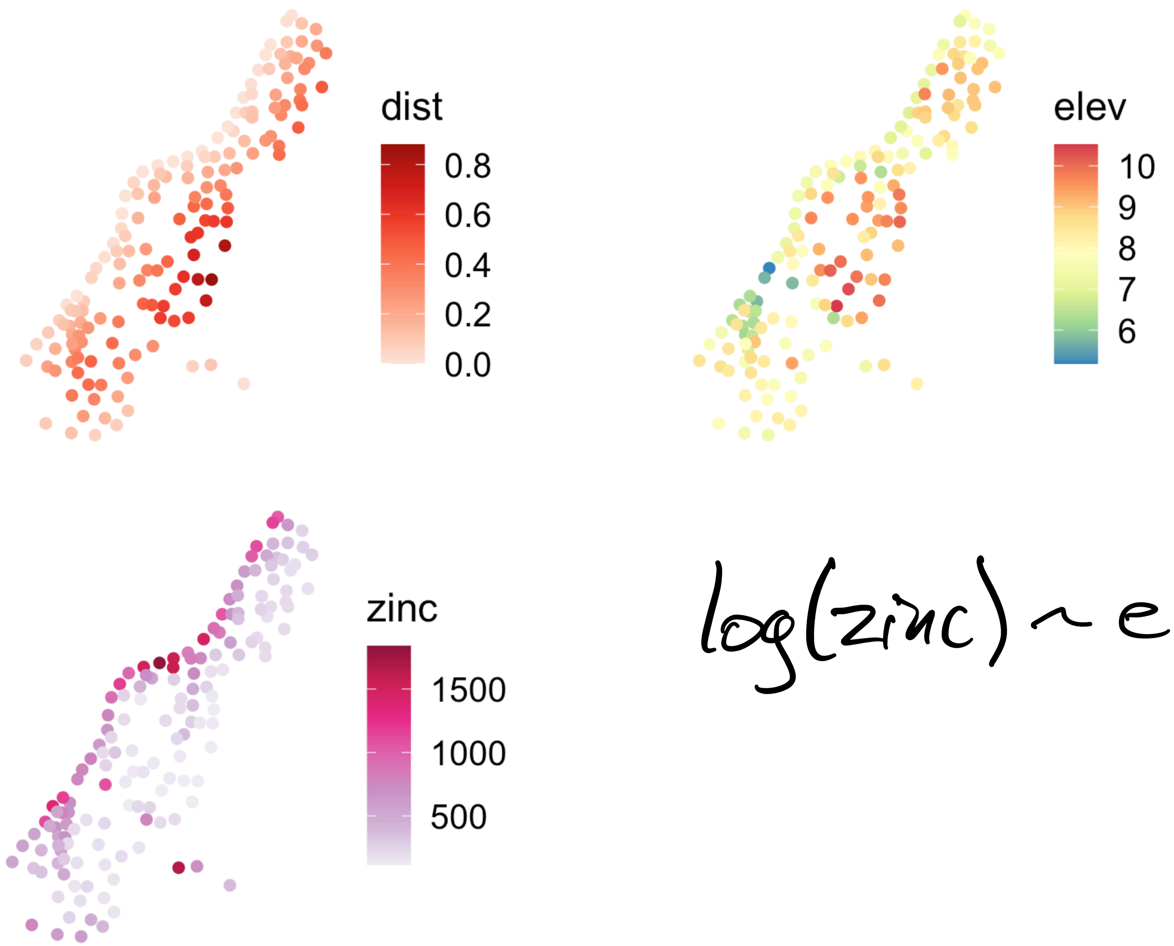
Error models

Unobserved variable is introducing autocorrelated residuals



Solutions

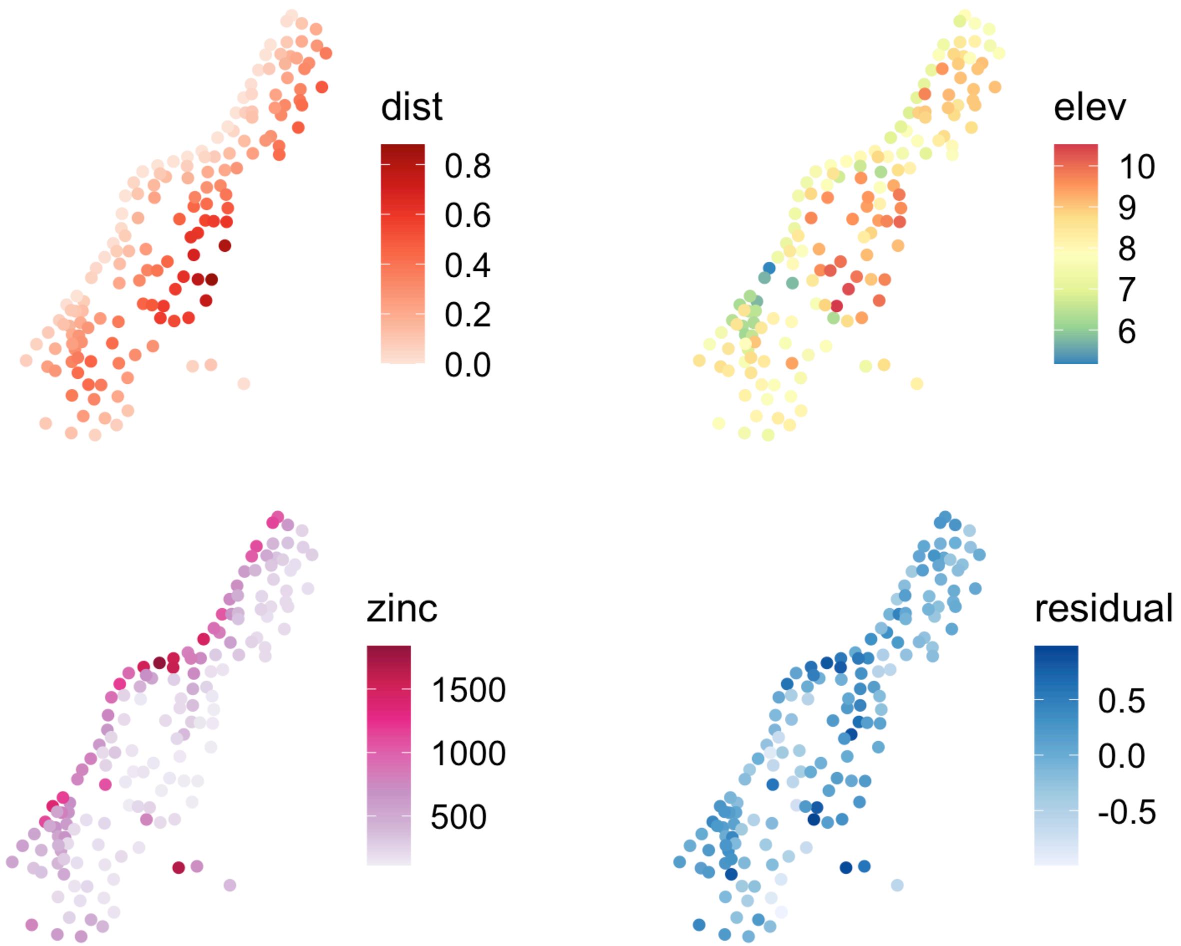
Error models



log(zinc) ~ elev + sqrt(dist)

Solutions

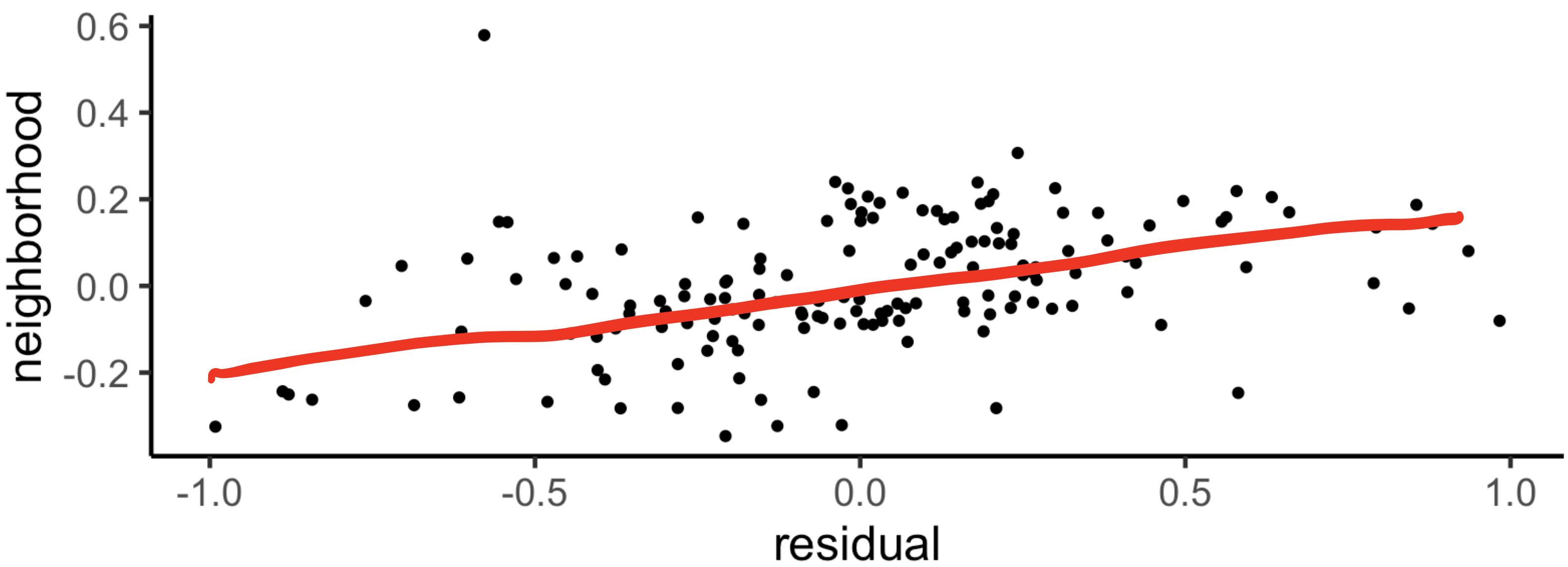
Error models



Solutions

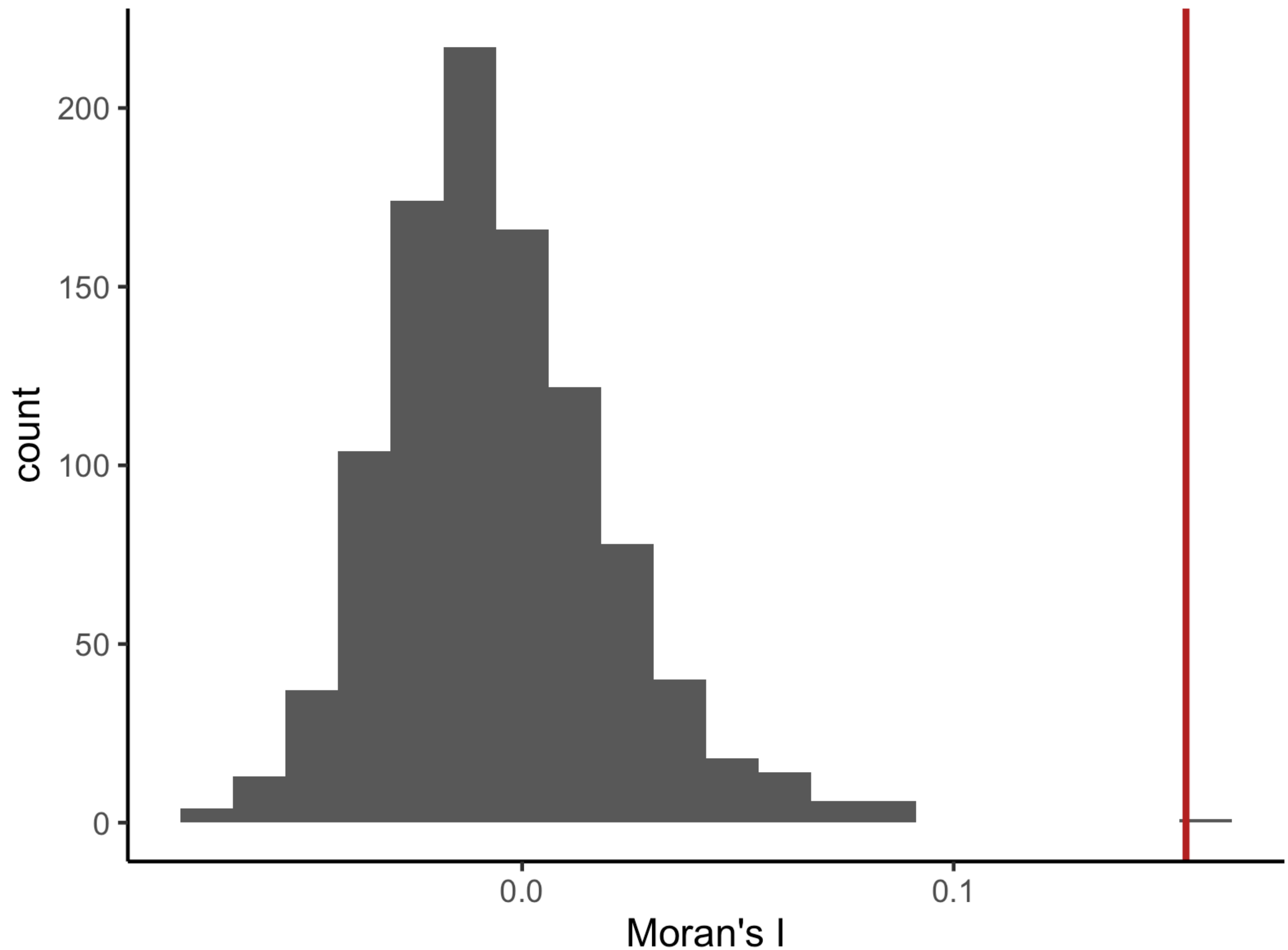
Error models

```
ols_mod <- lm(log(zinc) ~ elev + sqrt(dist), data = meuse)
meuse$residual <- resid(ols_mod)
meuse.nb <- dnearneigh(meuse_sf, d1 = 0, d2 = 500)
meuse.lw <- nb2listw(meuse.nb, style = "W")
inc.lag <- lag.listw(meuse.lw, meuse_sf$residual)
tibble(residual = meuse$residual, neighborhood = inc.lag) %>%
  ggplot(aes(residual, neighborhood)) +
  geom_point()
```



Solutions

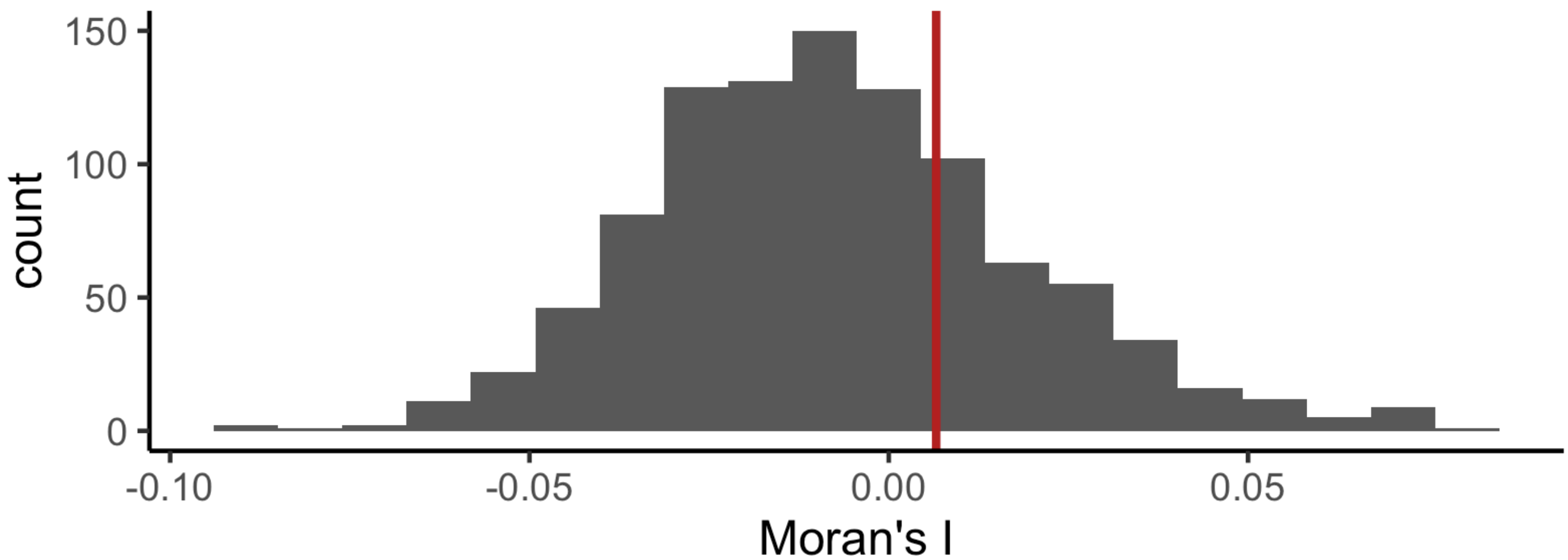
Error models



Solutions

Error models

```
meuse.nb <- dnearneigh(meuse_sf, d1 = 0, d2 = 500)
meuse.lw <- nb2listw(meuse.nb, style = "W")
sperr_mod <- errorsarlm(log(zinc) ~ elev + sqrt(dist),
                        data = meuse_sf,
                        listw = meuse.lw)
summary(sperr_mod)
```



Solutions

Error models

- Error models incorporate autocorrelation into the error term
- Works if an unobserved variables is causing autocorrelation
- Requires you to define the neighborhood - and can be sensitive to this choice
- Also applicable to temporal models!

Summary

- **Autocorrelation**
  - Things close in time or space tend to be similar
  - Autocorrelated residuals are a big no-no; use tests to identify them
- **Solutions**
  - Lag models incorporate information from nearby observations - great when past variables influence their future values
  - Error models account for autocorrelation directly in the residuals - useful for unobserved sources of autocorrelation