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

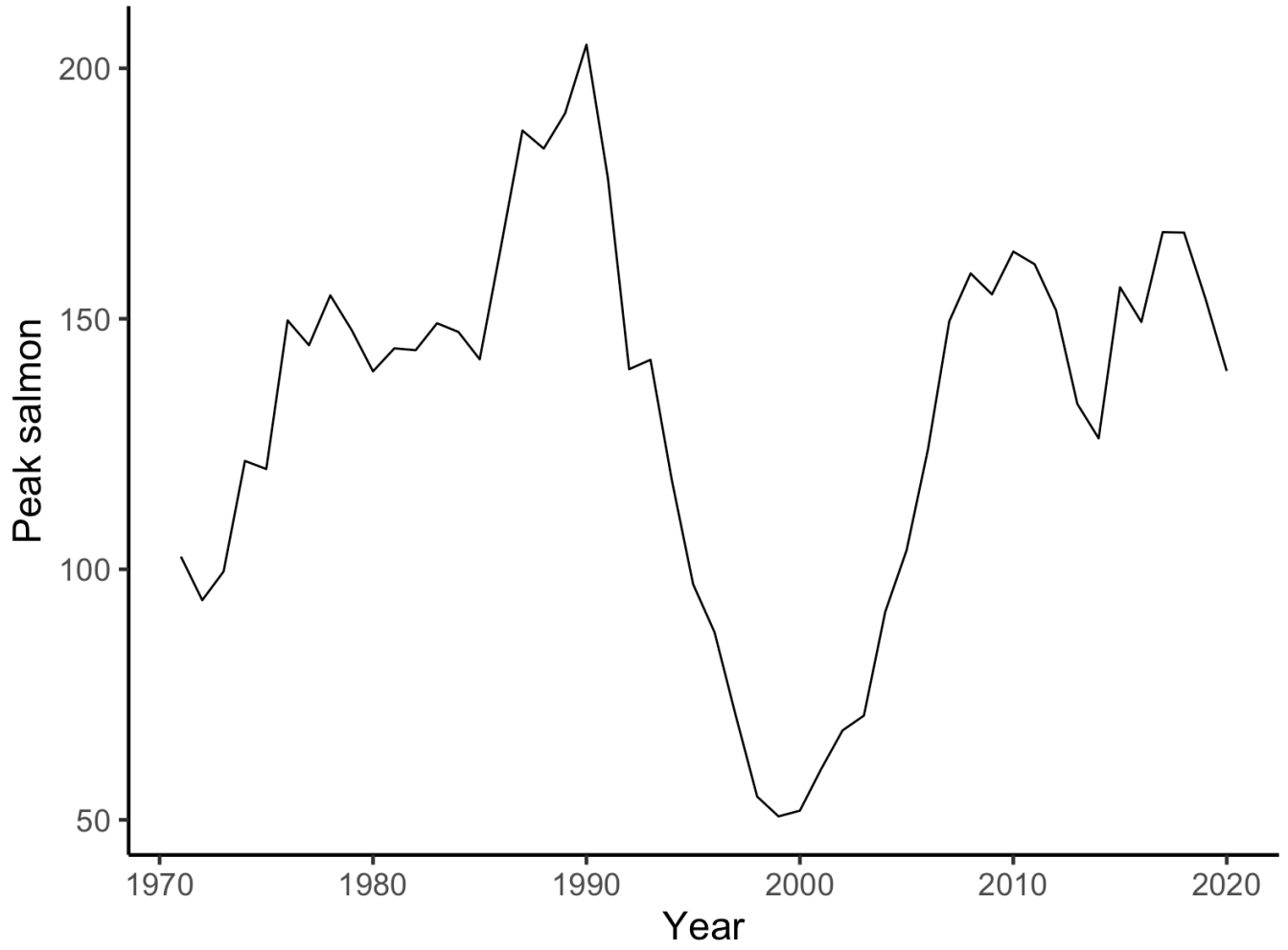
What is it?

Autocorrelation

What is it?

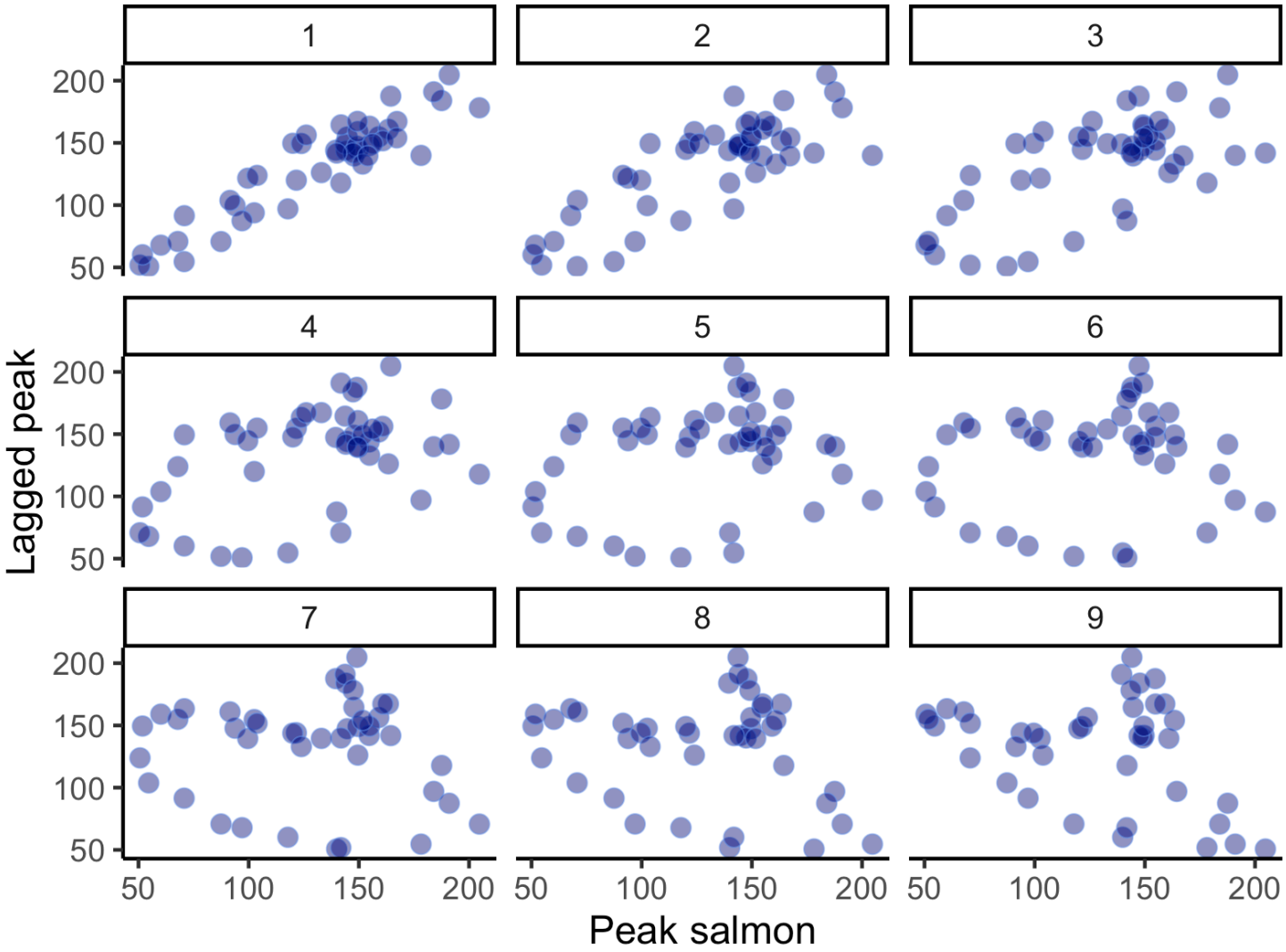
Autocorrelation

What is it?



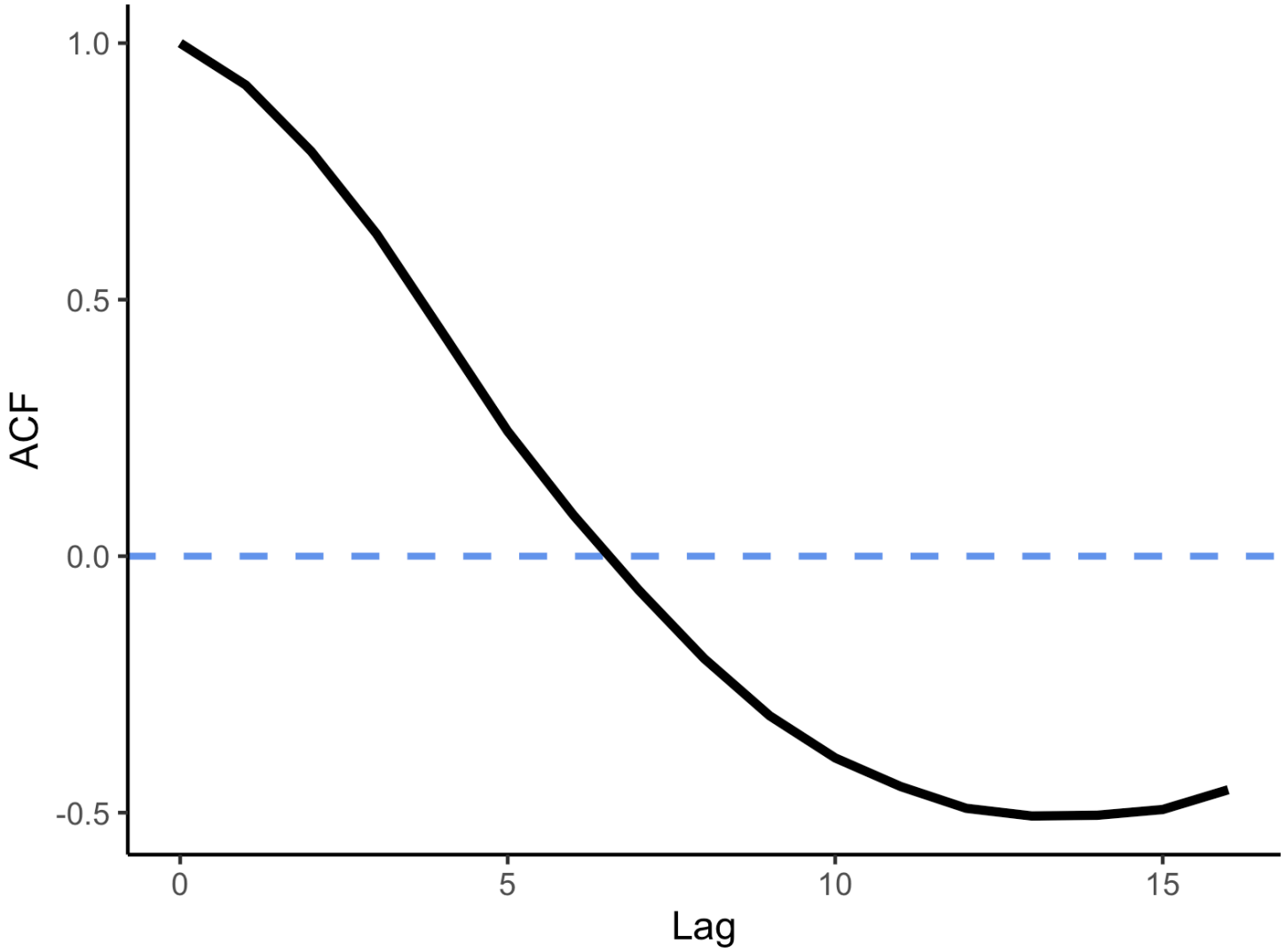
Autocorrelation

What is it?



Autocorrelation

What is it?



Autocorrelation

What is it?

Autocorrelation

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

Autocorrelation

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

Autocorrelation

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

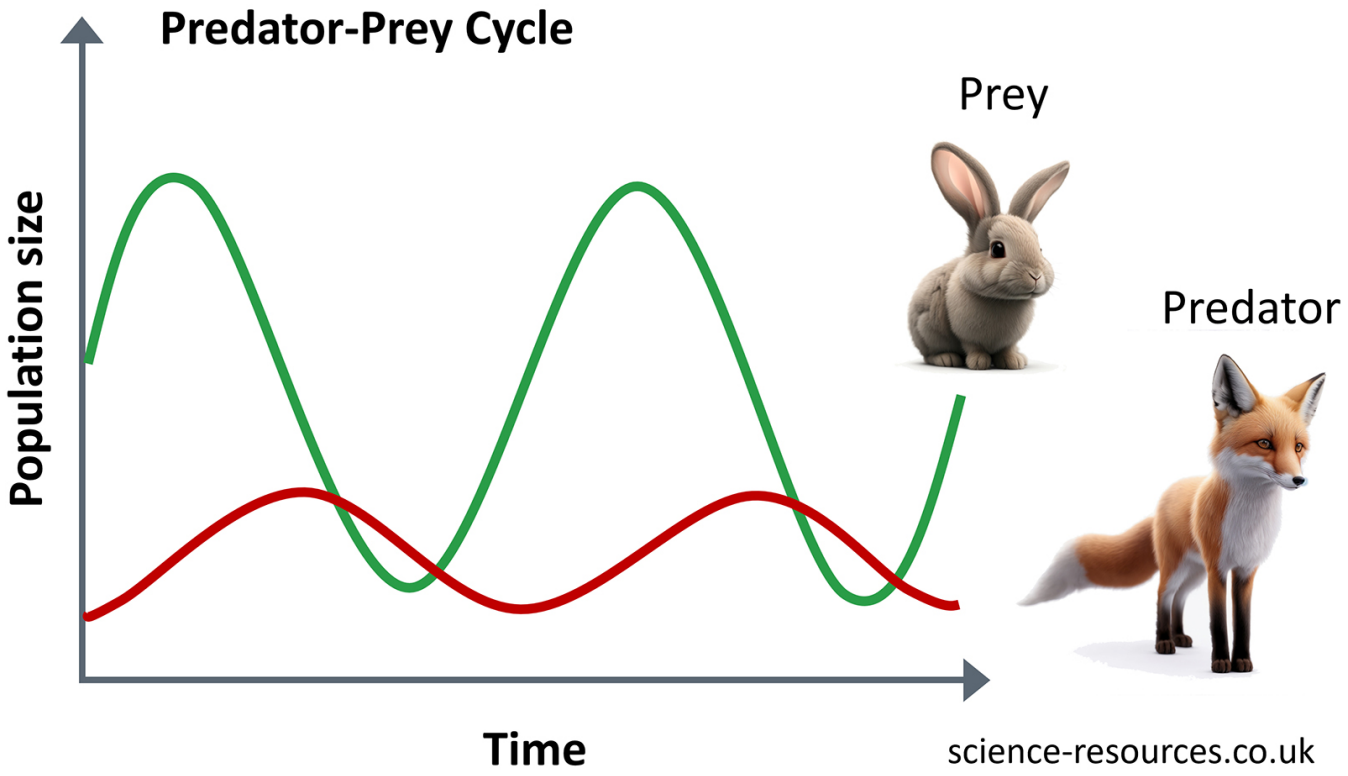
Autocorrelation

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

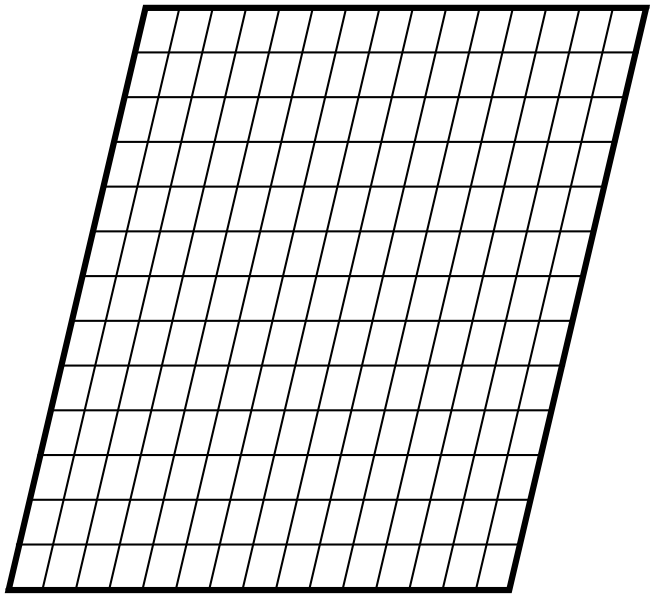
Solutions

Depends on nature of autocorrelation



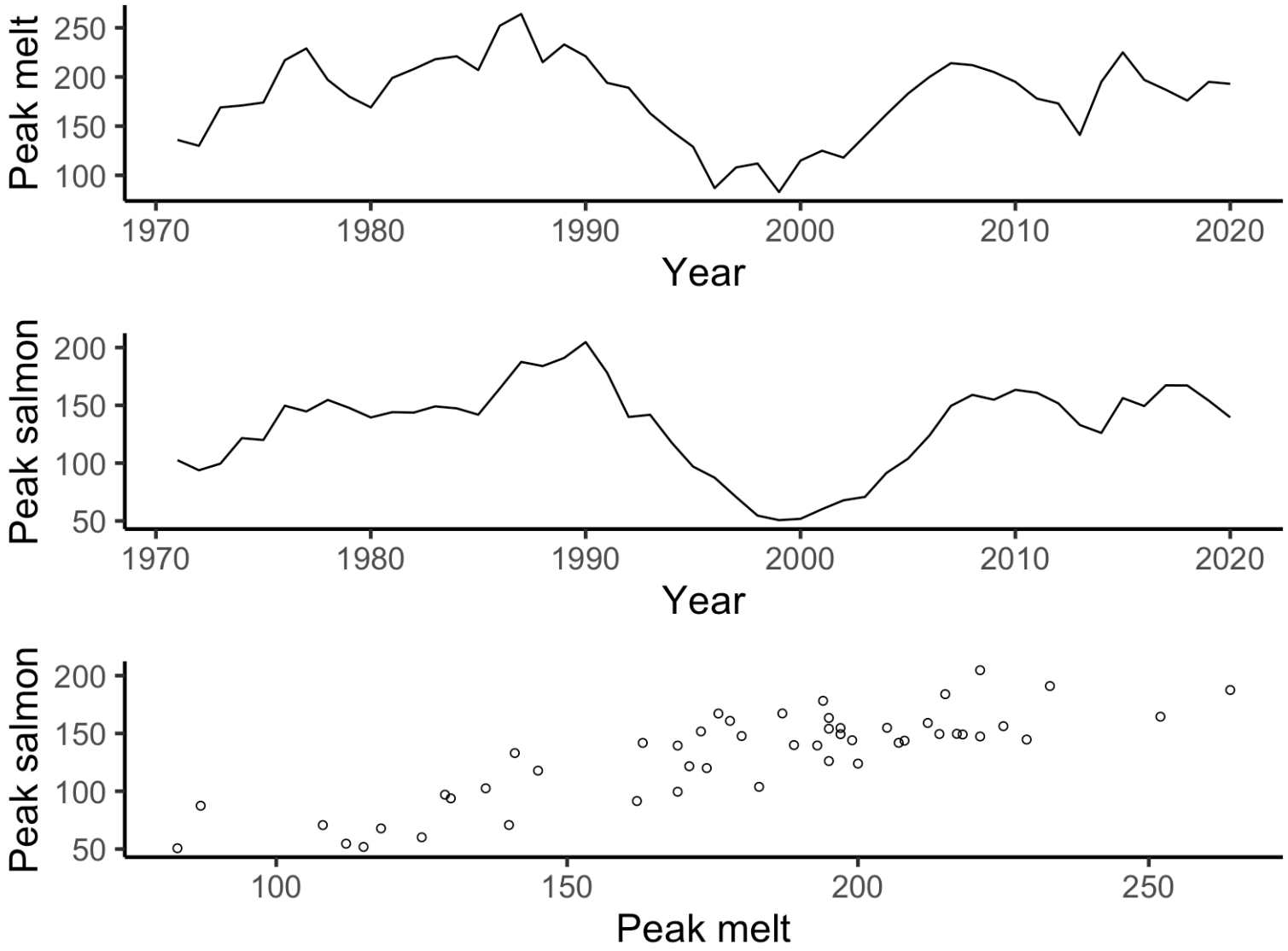
Solutions

Depends on nature of autocorrelation



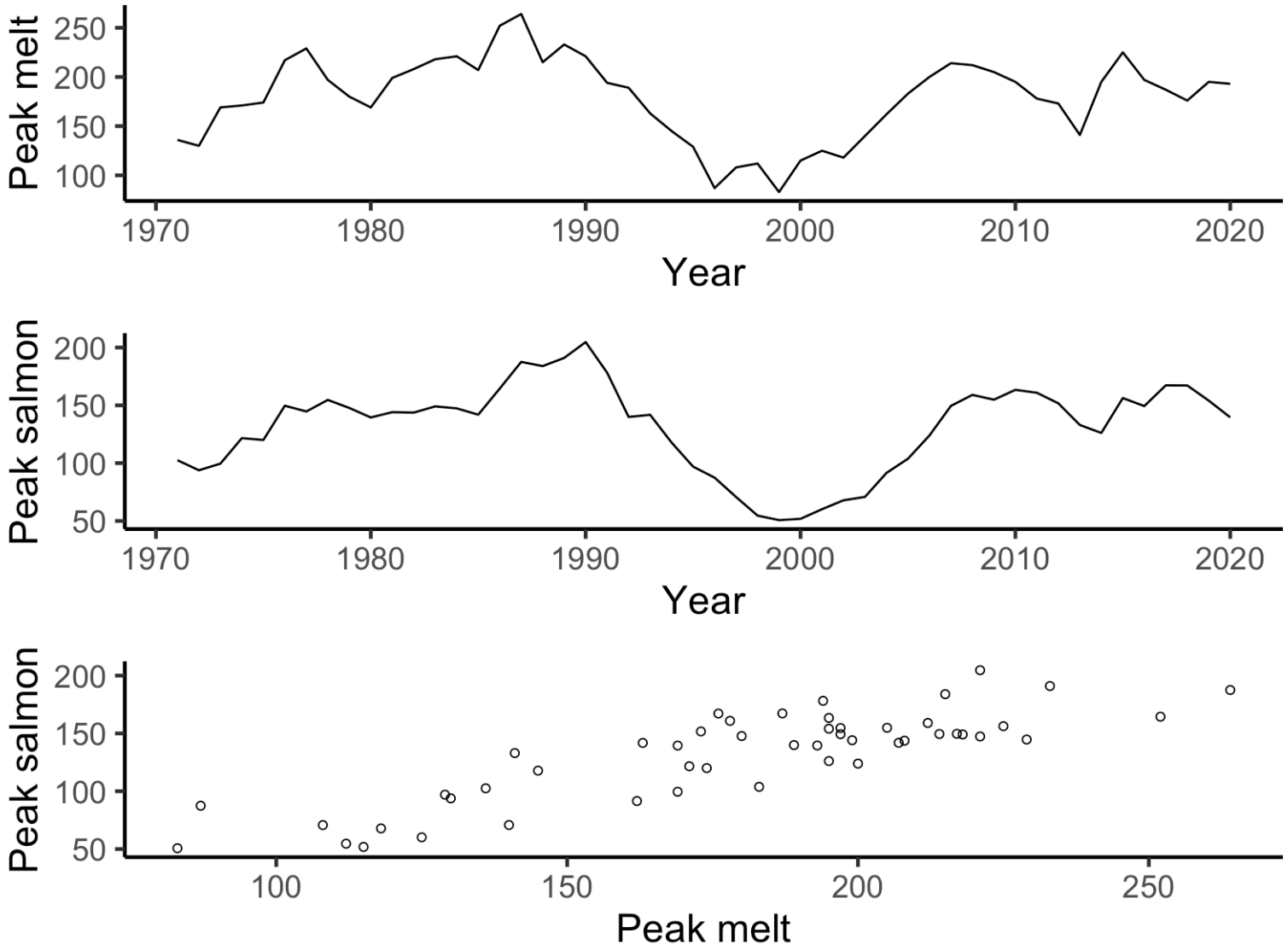
Solutions

Lag models



Solutions

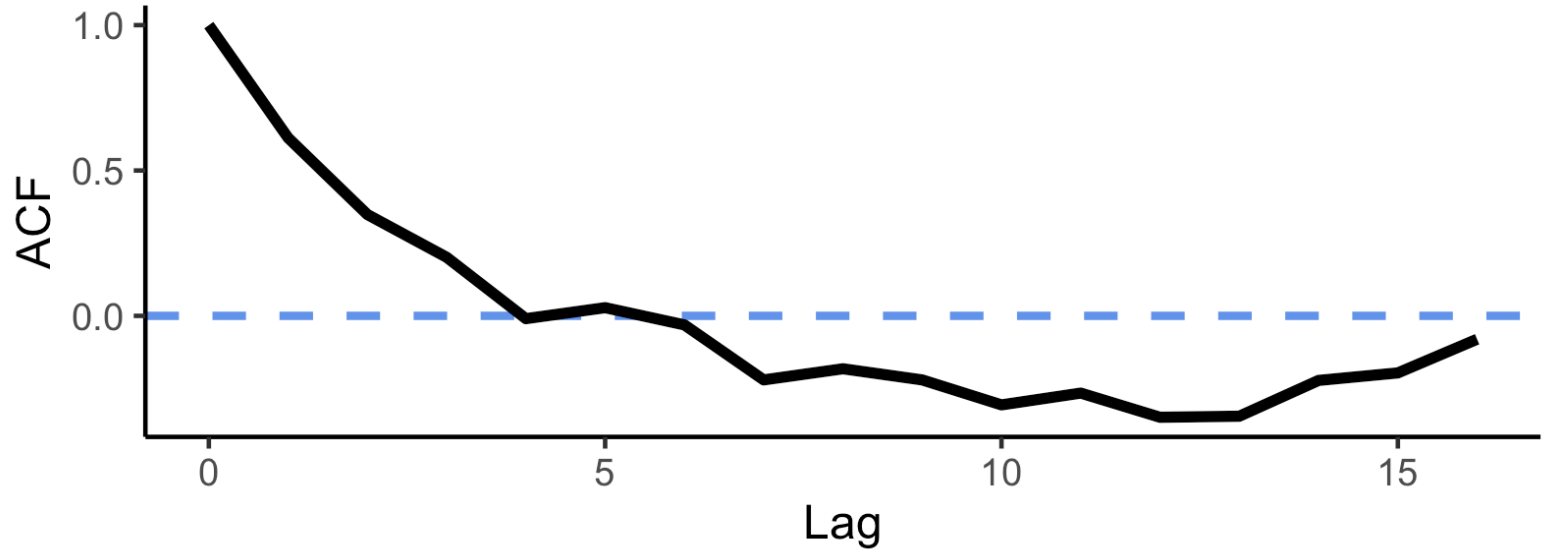
Lag models



Solutions

Lag models

```
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)
```



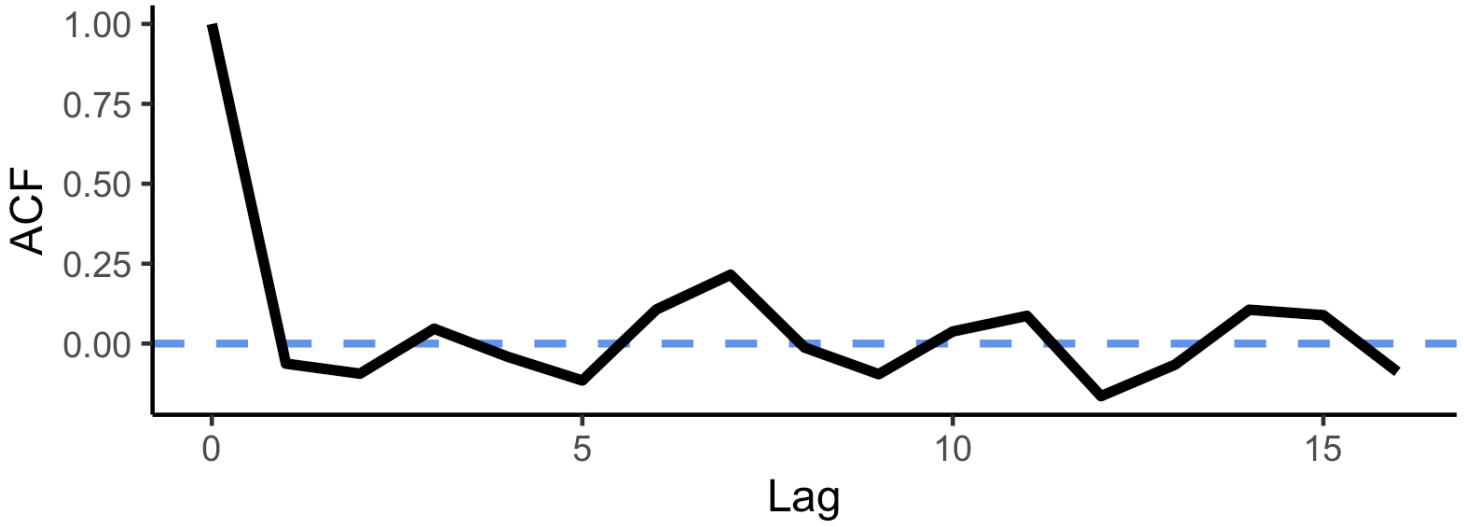
Solutions

Lag models

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)
```



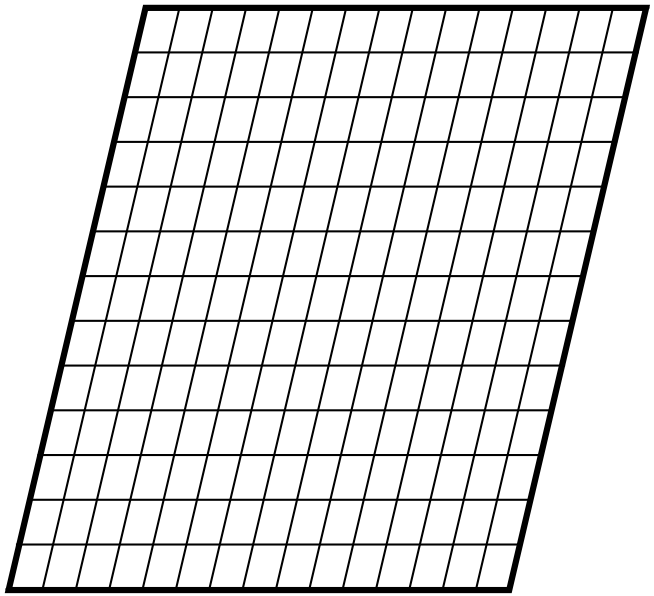
Solutions

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!

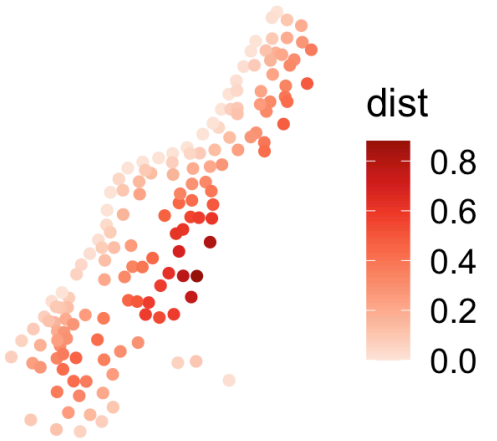
Solutions

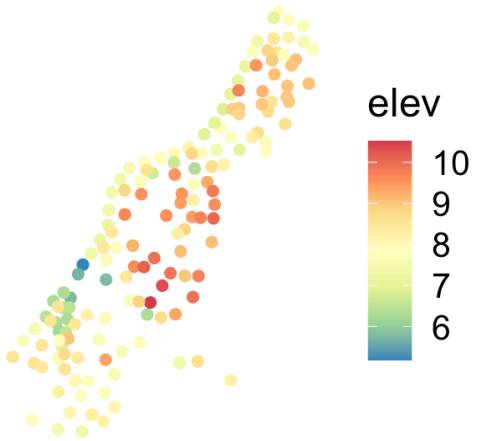
Error models

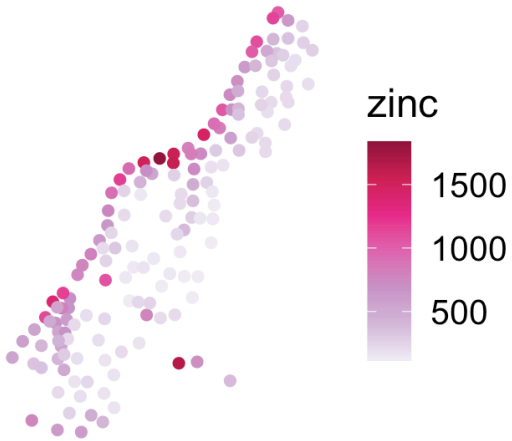


Solutions

Error models

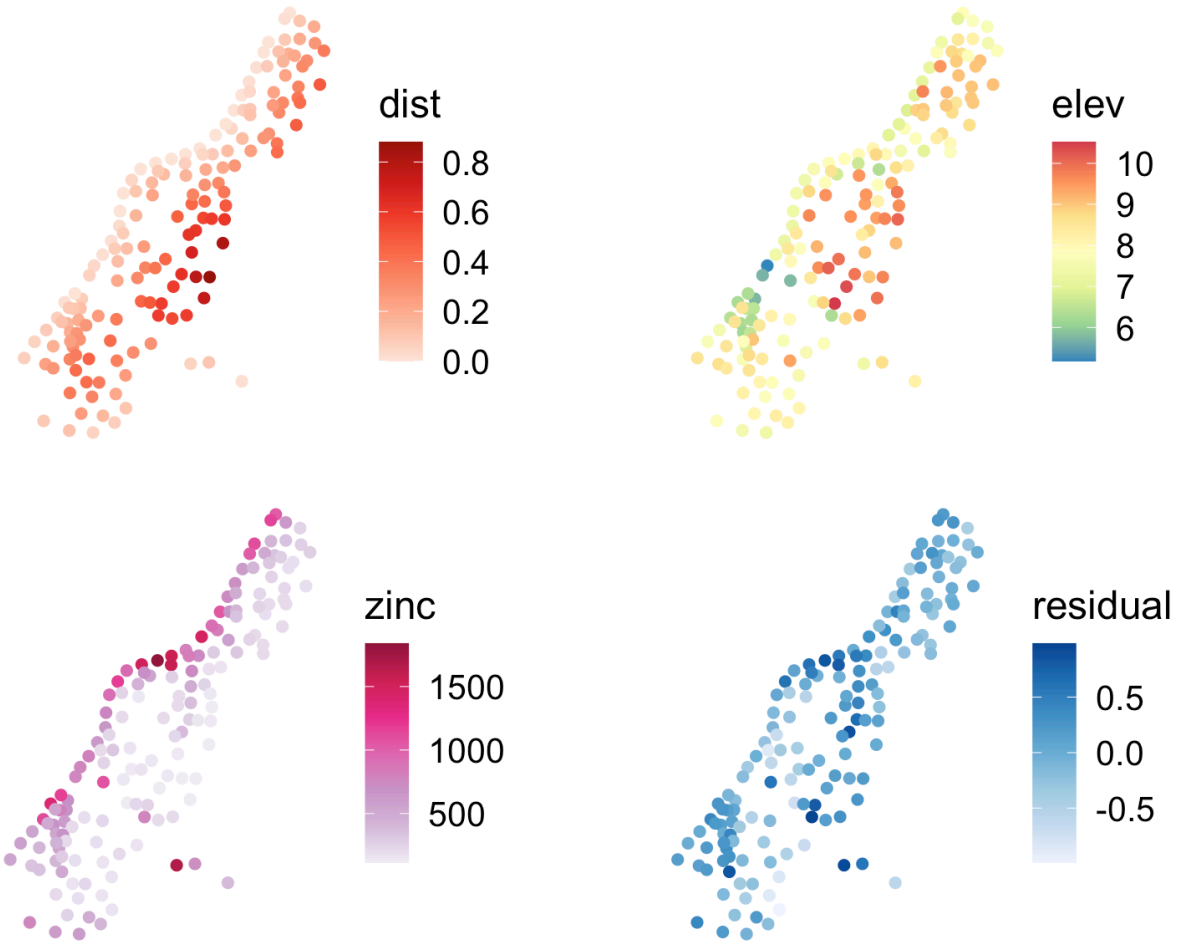






Solutions

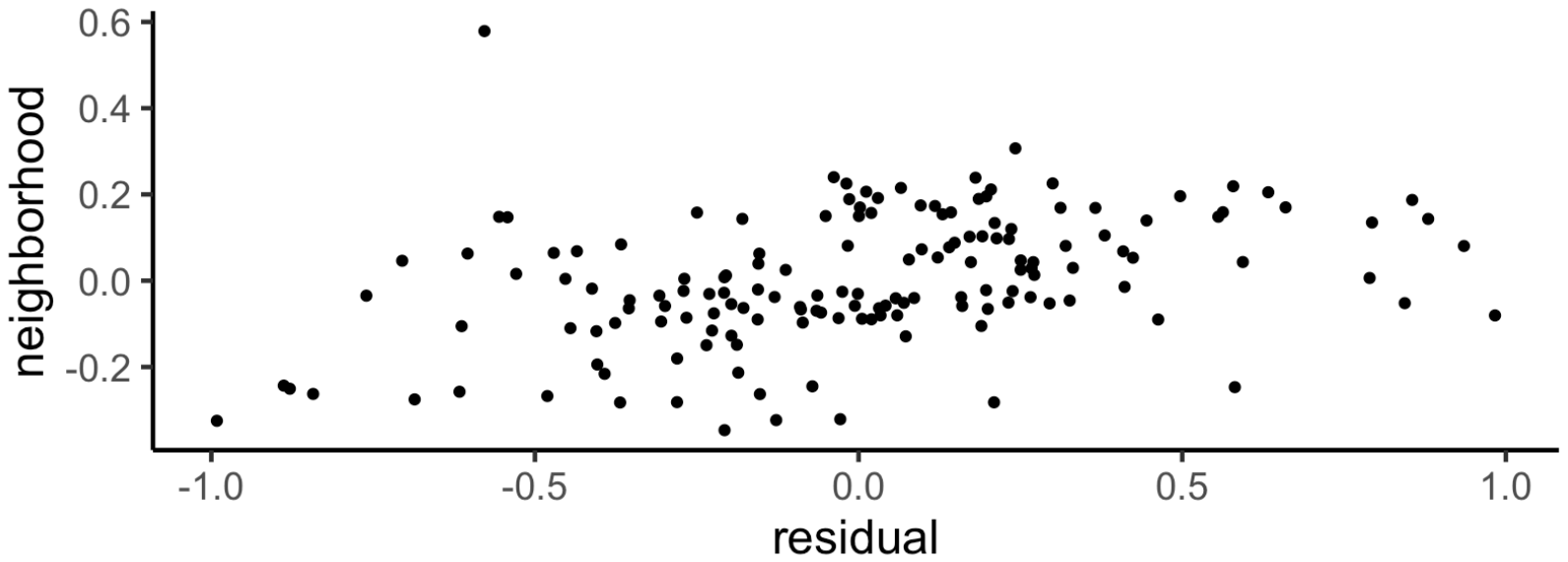
Error models



Solutions

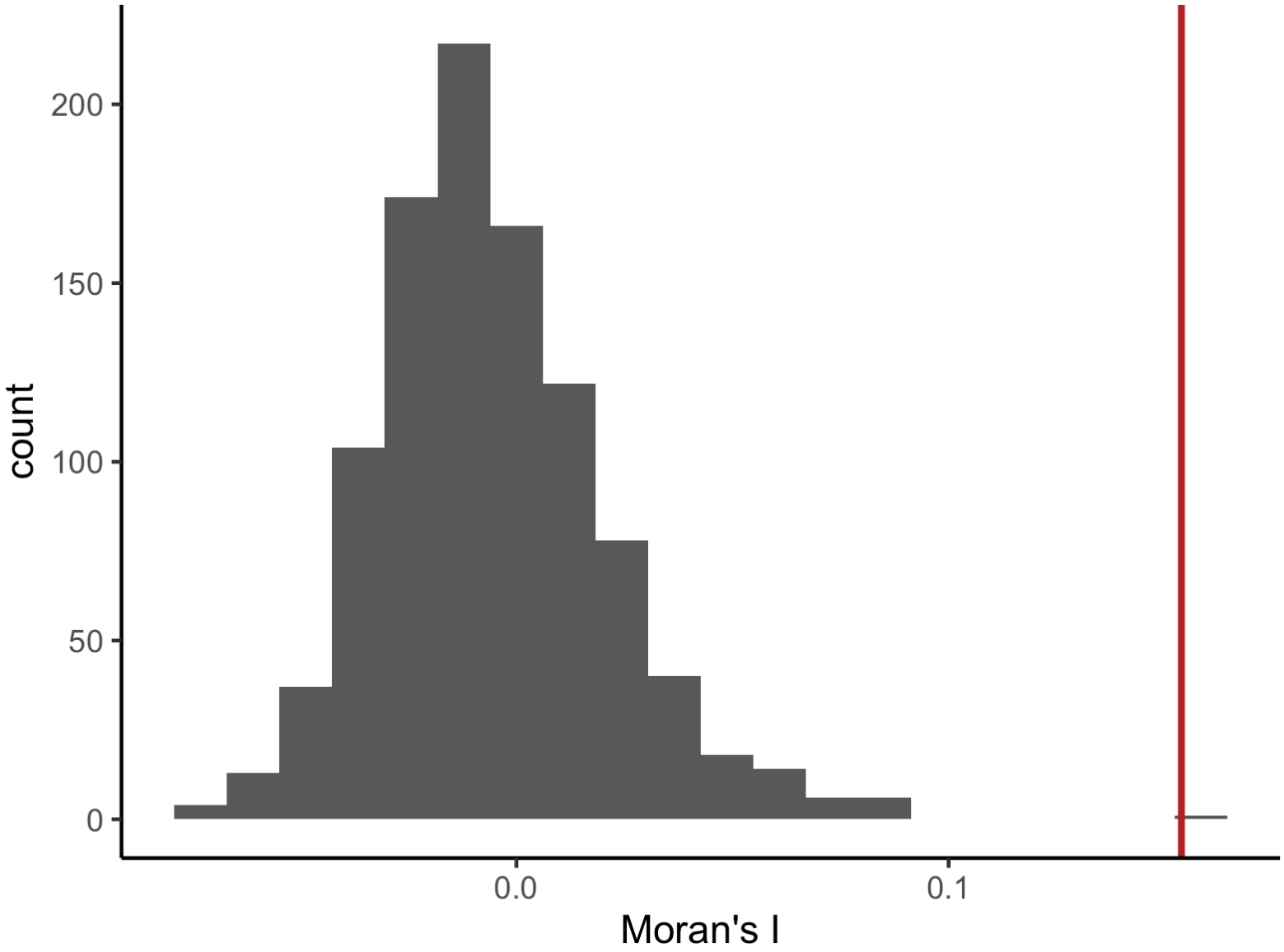
Error models

```
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

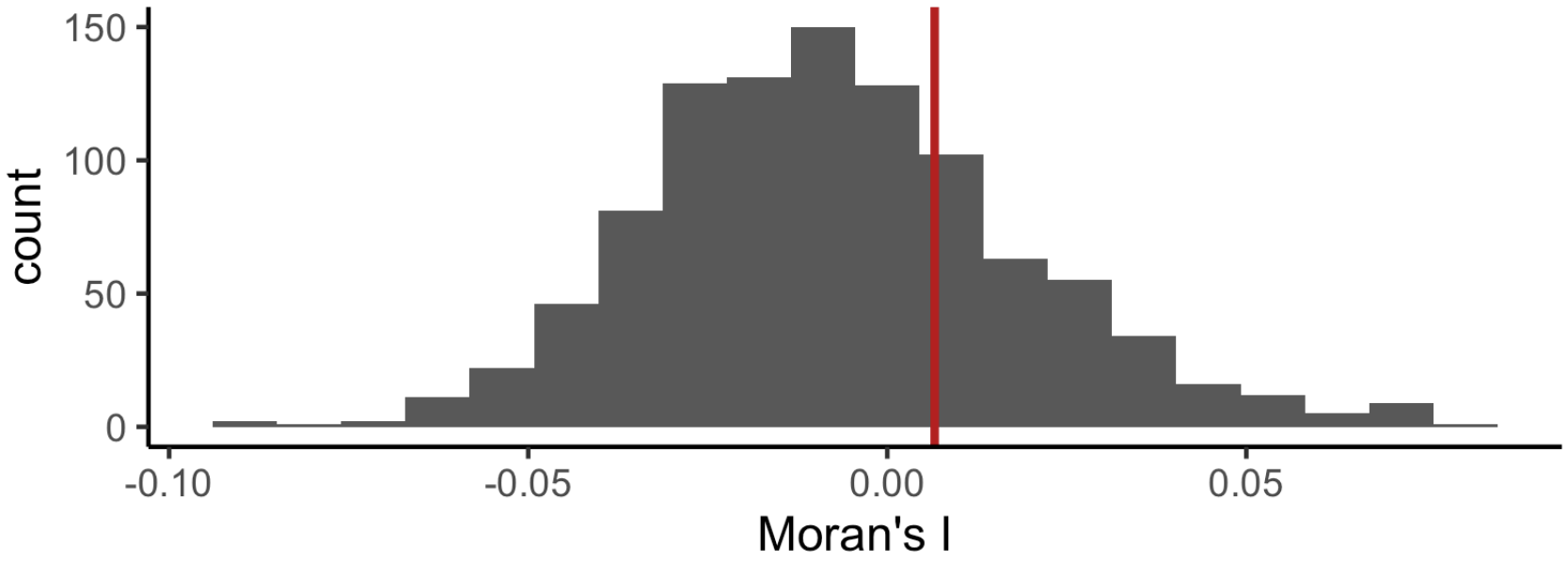
Error models



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