ITSx: Policy Analysis Using Interrupted Time Series

Week 2 Slides

Michael Law, Ph.D.
The University of British Columbia

Layout of the weeks

1. Introduction, setup, data sources
2. $\boxed{\text{Single series interrupted time series analysis}}$
3. ITS with a control group
4. Extensions and regression discontinuities
5. Course wrap-up

Overview of steps

1. Determine time periods
2. Select analytic cohorts
3. Determine outcomes of interest
4. Setup data
5. Visually inspect the data
6. Perform preliminary analysis
7. Check for and address autocorrelation
8. Run the final model
9. Plot the results
10. Predict relative and absolute effects

INTRODUCTION TO THE EXAMPLES

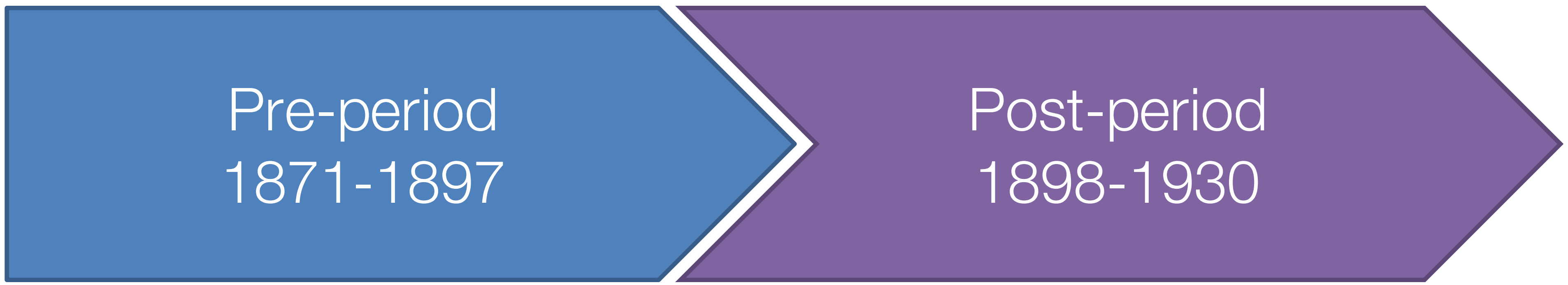
EXAMPLE 1: WATER FLOW ON NILE

Source: http://commons.wikimedia.org/wiki/File:River_Nile_map.svg


Source: http://commons.wikimedia.org/wiki/File:Aswan_low_dam2.JPG

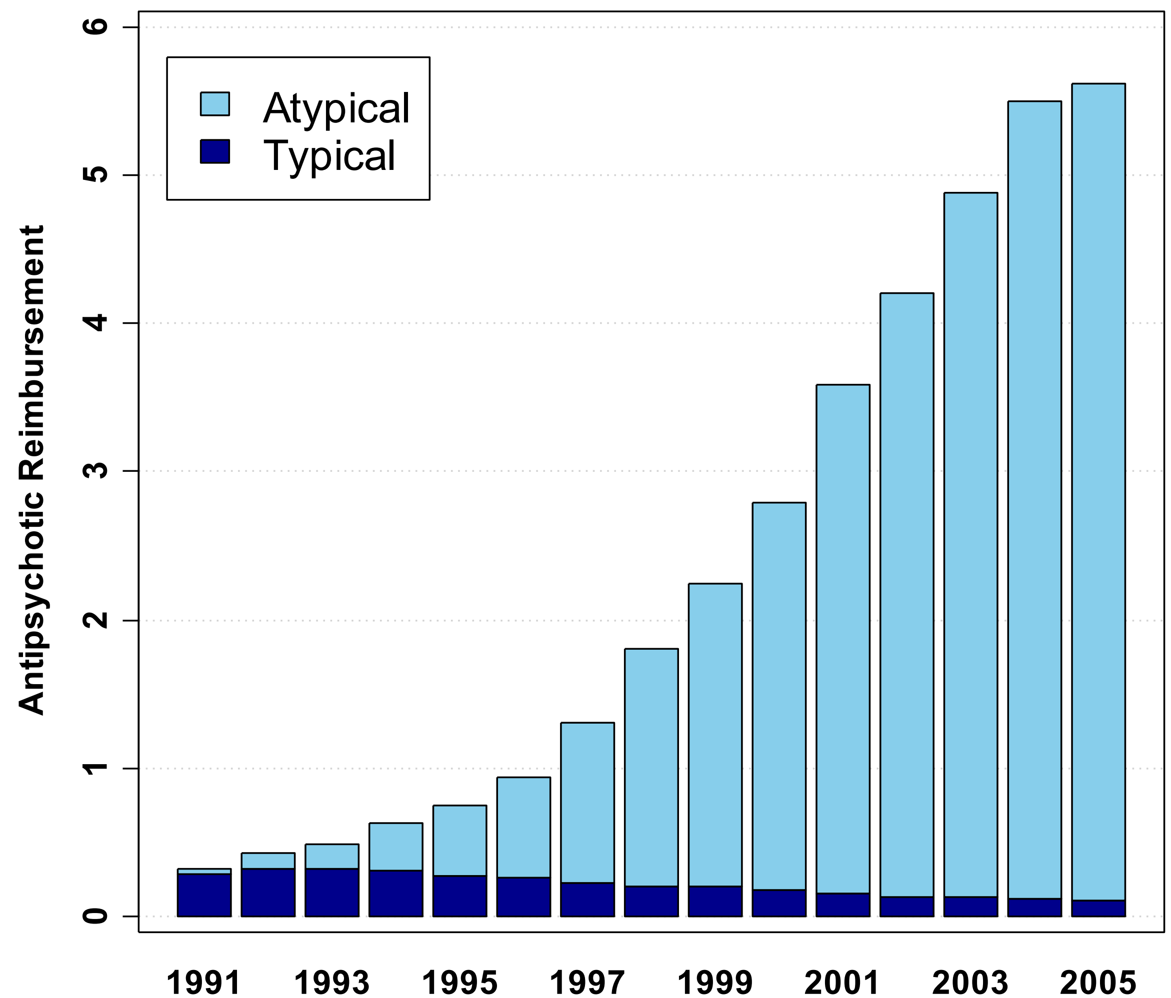
Research Question

- Previous work has suggested weather patterns shifted in 1898 (Cobb 1978)
- What was the impact of weather changes on annual water flow levels in the Nile?


Pre-period
1871-1897
Post-period
1898-1930

EXAMPLE 2: WEST VIRGINIA MEDICAID DRUG POLICY

Medicaid Antipsychotic Reimbursement, 1991-2005



Source: CMS Medicaid Quarterly Drug Utilization Data, excludes Arizona, which does not provide data
Converted to 2005$ using the Medical Care component of the CPI

General Drug Prior Authorization Form

West Virginia Medicaid
Drug Prior Authorization Form
http://www.dhhr.wv.gov/bms/Pharmacy/Pages/default.aspx

Rational Drug Therapy Program
WVU School of Pharmacy
PO Box 9511 HSCN
Morgantown, WV 26506
Fax: 1-800-531-7787
Phone: 1-800-847-3859

Patient Name (Last) | (First) | (M) | WV Medicaid 11 Digit ID# | Date of Birth (MM/DD/YYYY)

Prescriber Name (Last) | (First) | (MI)

Prescriber Address (Street) | (City) | (State) West Virginia | (Zip)

Prescriber 10-Digit NPI# | Phone # (111-222-3333) | Fax # (111-222-3333)

Pharmacy Name (if applicable)

Pharmacy Address (Street) | (City) | (State) West Virginia | (Zip)

Pharmacy 10-Digit NPI# | Phone # (111-222-3333) | Fax # (111-222-3333)

state	yearqtr	policy	marketshare
WV	2001, 2	0	0.527
WV	2001, 3	0	0.5
WV	2001, 4	0	0.492
WV	2002, 1	0	0.497
WV	2002, 2	0	0.497
WV	2002, 3	0	0.497
WV	2002, 4	0	0.5
WV	2003, 1	0	0.501
WV	2003, 2	1	0.464
WV	2003, 3	1	0.434
WV	2003, 4	1	0.423
WV	2004, 1	1	0.419
WV	2004, 2	1	0.386
WV	2004, 3	1	0.37
WV	2004, 4	1	0.357
WV	2005, 1	1	0.338
WV	2005, 2	1	0.311
WV	2005, 3	1	0.305
WV	2005, 4	1	0.289

Source: Law MR, Ross-Degnan D, Soumerai SB. Effect of Prior Authorization of Second-Generation Antipsychotic Agents on Pharmacy Utilization and Reimbursements. Psychiatric Services 2008 59:5, 540-546

STEP 4: SETUP DATA

Data Setup

- One data row for each time period
- Variables:
  - Existing Trend
  - Post-intervention Level Change
  - Post-intervention Trend Change
  - Outcome of interest
- Data Source: Vorosmarty, C. J., B. M. Fekete, and B. A. Tucker. 1998. Global River Discharge, 1807-1991, Version. 1.1 (RivDIS). Data set. Available on-line [http://www.daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

year	flow
1871	3533.9
1872	3664.3
…	…
1896	3856.0
1897	3248.2
1898	3479.8
1899	2453.0
…	…
1929	3241.1
1930	2334.4

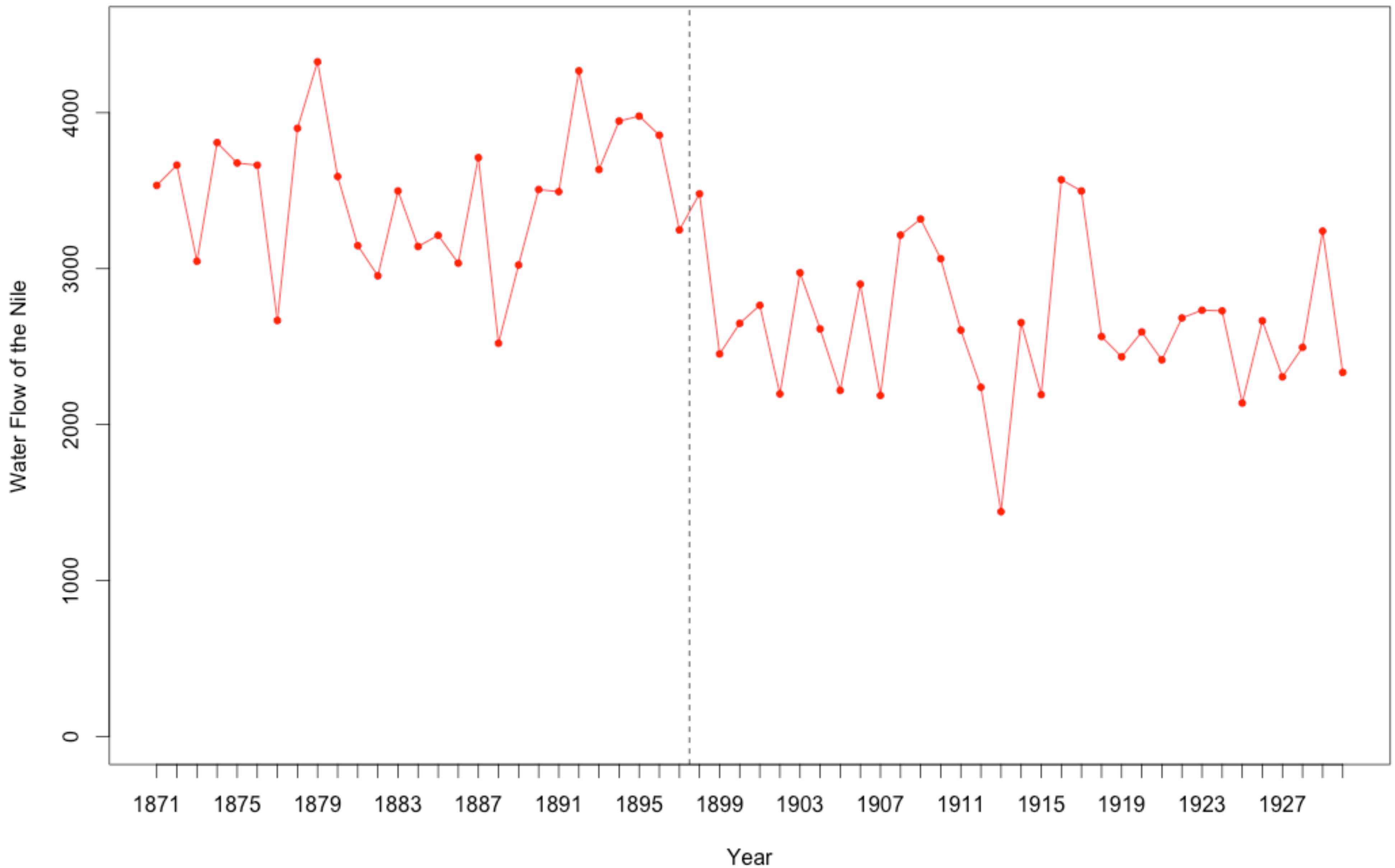
year	time	flow	level	trend
1871	1	3533.9	0	0
1872	2	3664.3	0	0
…	…	…	…	…
1896	26	3856.0	0	0
1897	27	3248.2	0	0
1898	28	3479.8	1	1
1899	29	2453.0	1	2
…	…	…	…	…
1929	59	3241.1	1	32
1930	60	2334.4	1	33

STEP 5: VISUALLY INSPECT DATA (2)

Visually Inspect Data

- What you are looking for:
  - “Wild” points
  - Linear trends
  - Co-interventions
  - Data quality issues

```
# Plot outcome variable versus time
plot(data$time,data$flow,
       ylab="Water Flow of the Nile",
       ylim=c(0,4500),
       xlab="Year",
       type="l",
       col="red",
       xaxt="n")

# Add x-axis year labels
axis(1, at=1:60, labels=data$year)

# Add in the points for the figure
points(data$time,data$flow,
       col="red",
       pch=20)

# Label the weather change
abline(v=27.5,lty=2)
```



Water Flow of the Nile
0
1000
2000
3000
4000
1871
1875
1879
1883
1887
1891
1895
1899
1903
1907
1911
1915
1919
1923
1927
Year

STEP 6: PRELIMINARY ANALYSIS

Preliminary analysis

- First, run a standard OLS regression with a time series specification

- This will form the basis for checks about autocorrelation

Basic time series model

- For intervention status *j*, at time *t:*

$$outcome_{jt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot level_j + \beta_3 \cdot trend_{jt} + \varepsilon_{jt}$$

OLS Regression in R

```
# A preliminary OLS regression
model_ols <- lm(flow ~ time + level + trend,
data=data)

# See summary of model output
summary(model_ols)

# Get confidence intervals for coefficients
confint(model_ols)
```

OLS Model Results

```
Call:
lm(formula = flow ~ time + level + trend, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-1217.98  -309.60    -4.15   248.84   927.73

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3407.937    182.696  18.654  < 2e-16 ***
time           5.417     11.404   0.475  0.63659
level       -806.273    238.530  -3.380  0.00133 **
trend        -10.961     14.186  -0.773  0.44295
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 461.5 on 56 degrees of freedom
Multiple R-squared:  0.4651,  Adjusted R-squared:  0.4365
F-statistic: 16.23 on 3 and 56 DF,  p-value: 1.036e-07
```

STEP 7: AUTOCORRELATION

Overview of steps

1. Determine time periods
2. Select analytic cohorts
3. Determine outcomes of interest
4. Setup data
5. Visually inspect the data
6. Perform preliminary analysis
7. Check for and address autocorrelation
8. Run the final model
9. Plot the results
10. Predict relative and absolute effects

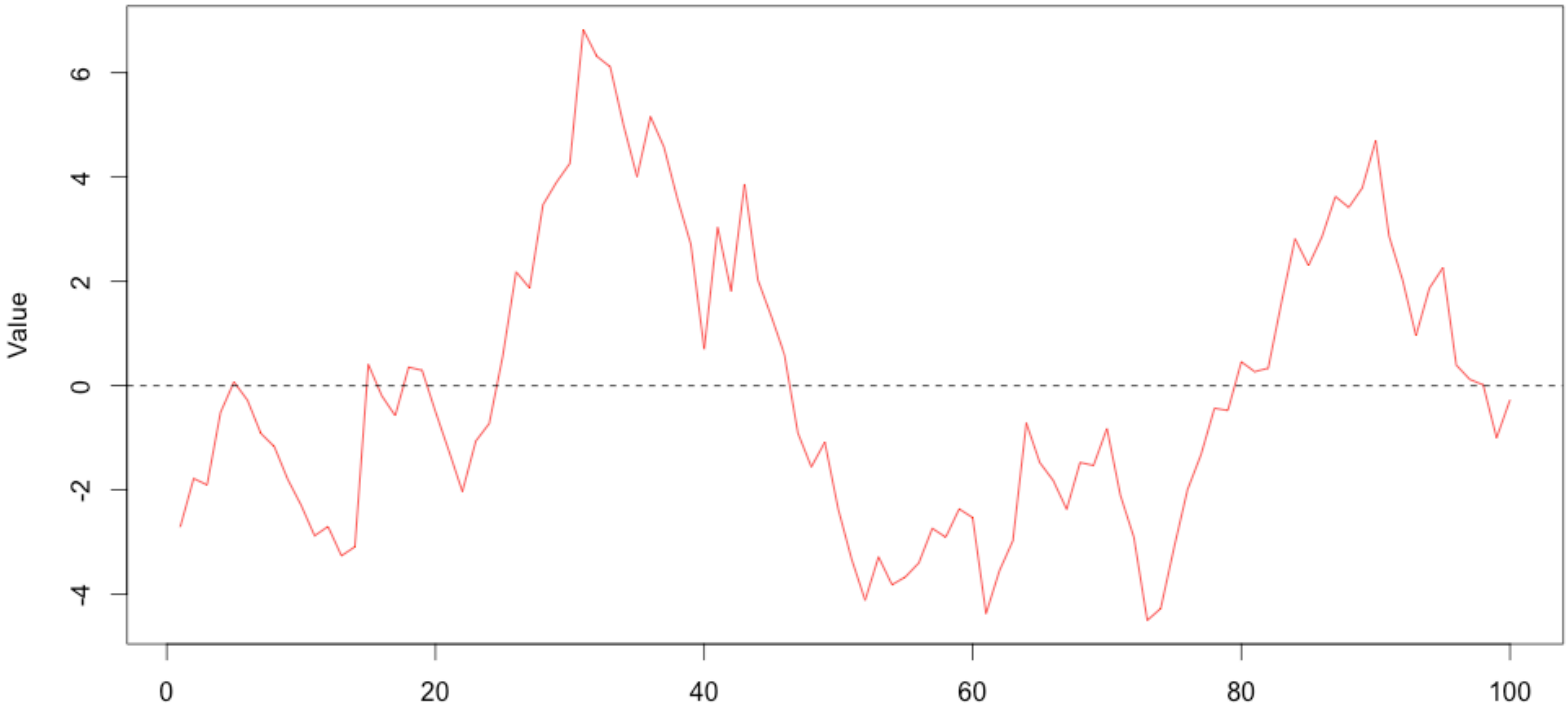
What is autocorrelation?

- Relationship in data points over time
  - Means the data points are not independent

- Two types we will discuss:
  - Autoregression
  - Moving Average

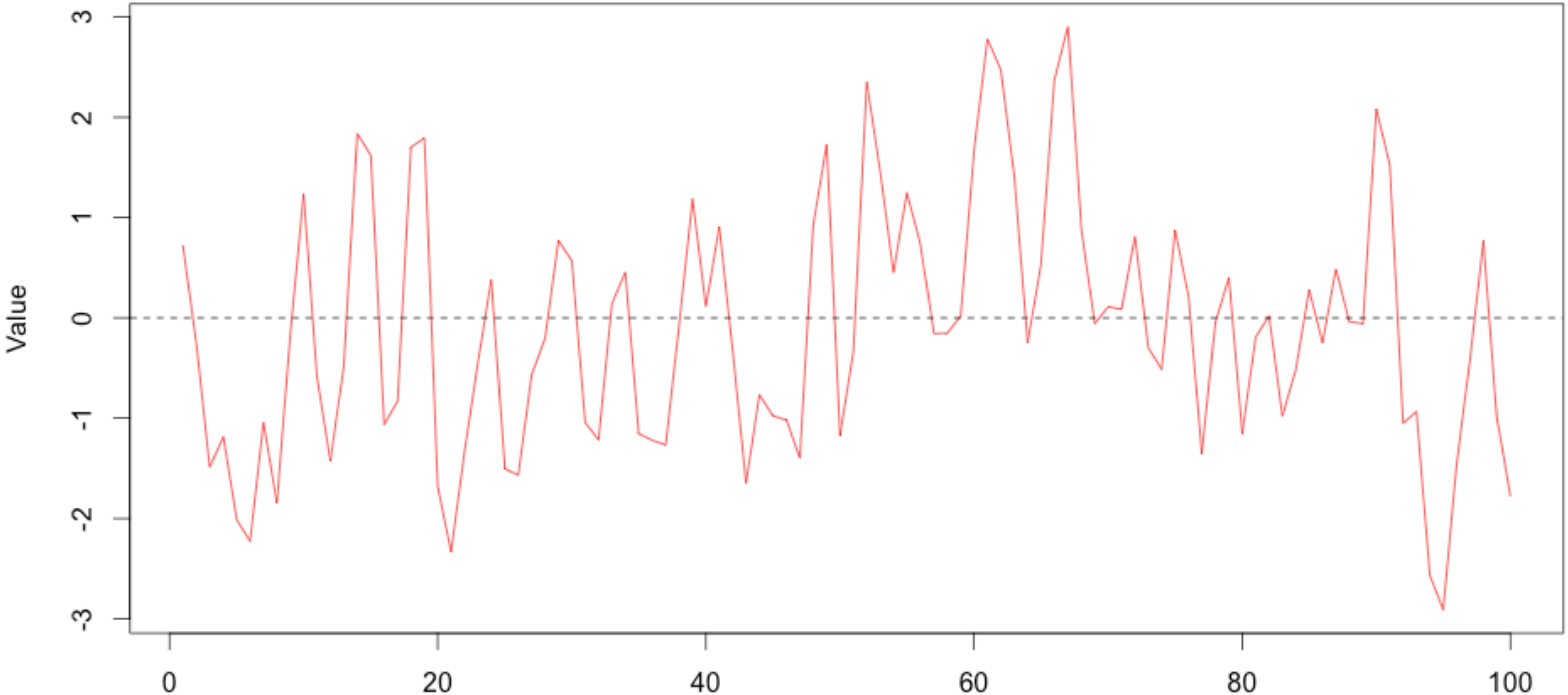
Autoregression

$$\varepsilon_t = \phi \varepsilon_{t-1} + \upsilon_t$$



Moving Average

$$\varepsilon_t = \upsilon_t + \vartheta \upsilon_{t-1}$$



Checking for Autocorrelation

- Check our earlier OLS regression

```
model_ols <- lm(flow ~ time + level +
   trend, data=data)
summary(model_ols)
```

Methods to Check

- Several methods, including:
  - Durbin-Watson test
  - Residual plots
  - ACF and partial-ACF plots

Durbin-Watson test

- A formal test that tests for correlated residuals
- Interpretation
  - Values of 2 indicate no autocorrelation
  - lower values indicate positive correlation, higher indicates negative correlation

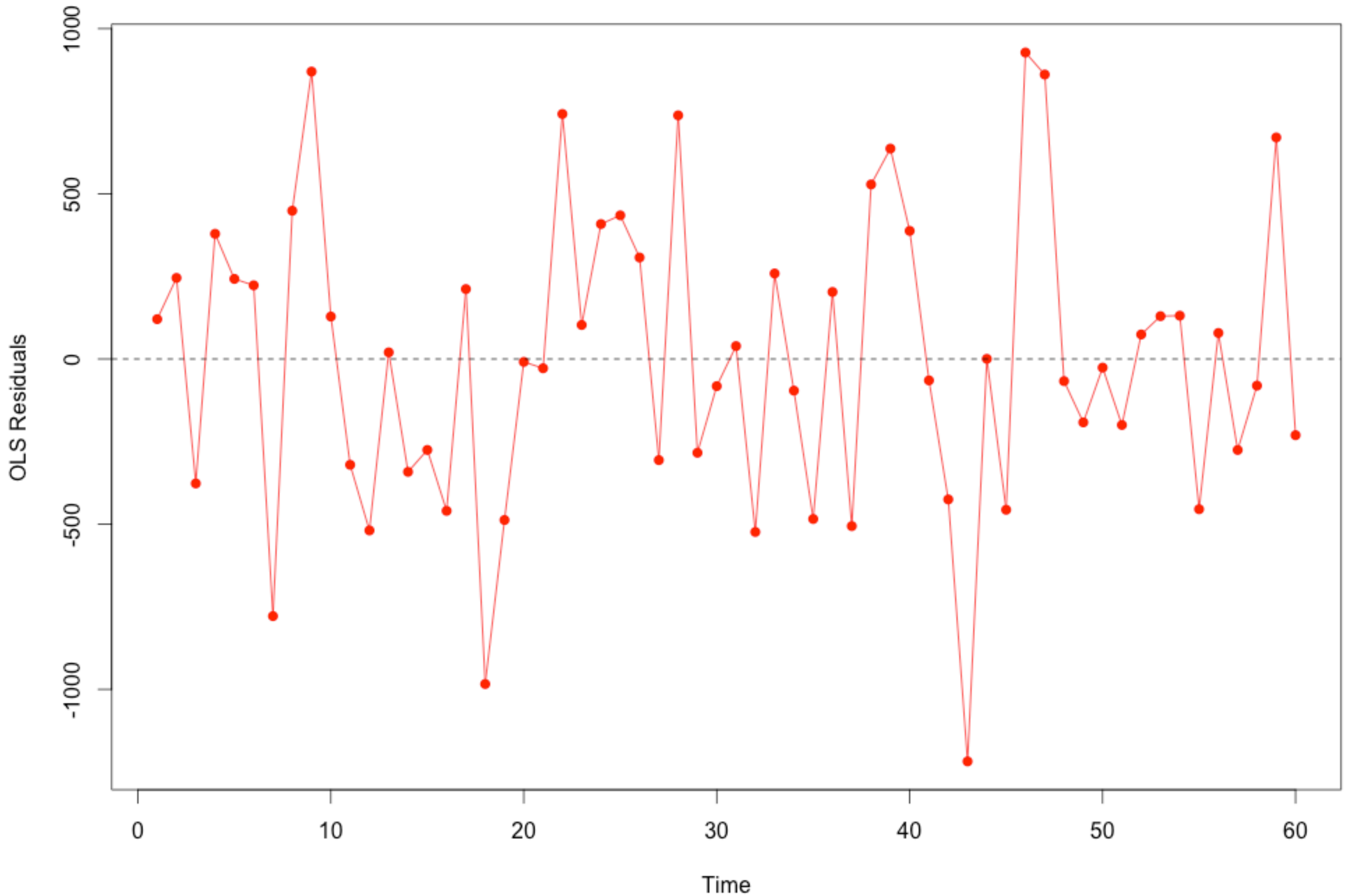
```
# Durbin-watson test, 12 time periods
dwt (model_ols,max.lag=12, alternative="two.sided")
```

```
> dwt(model_ols,max.lag=12,alternative="two.sided")
 lag Autocorrelation D-W Statistic p-value
   1    0.0734033049      1.847518   0.318
   2    0.0103495499      1.930880   0.648
   3   -0.1252424692      2.189623   0.492
   4   -0.2235224579      2.367786   0.128
   5   -0.0486230496      2.012538   0.758
   6    0.0021983903      1.889401   0.938
   7   -0.0268794370      1.895336   0.858
   8    0.1183775601      1.586528   0.394
   9   -0.1713776634      2.102128   0.196
  10   -0.3090629681      2.372771   0.008
  11    0.0268638508      1.692261   0.940
  12    0.0009121916      1.718511   0.740
 Alternative hypothesis: rho[lag] != 0
```

Residual plots

- Residuals from an OLS should not be related over time (independence assumption)
- Use a residual plot to visually inspect for patterns

```
# Plot residuals from OLS model
plot(data$time, residuals(ols),  type='o',
pch=16, xlab='Time', ylab='OLS Residuals')
```


OLS Residuals
Time
1000
500
0
-500
-1000
0
10
20
30
40
50
60

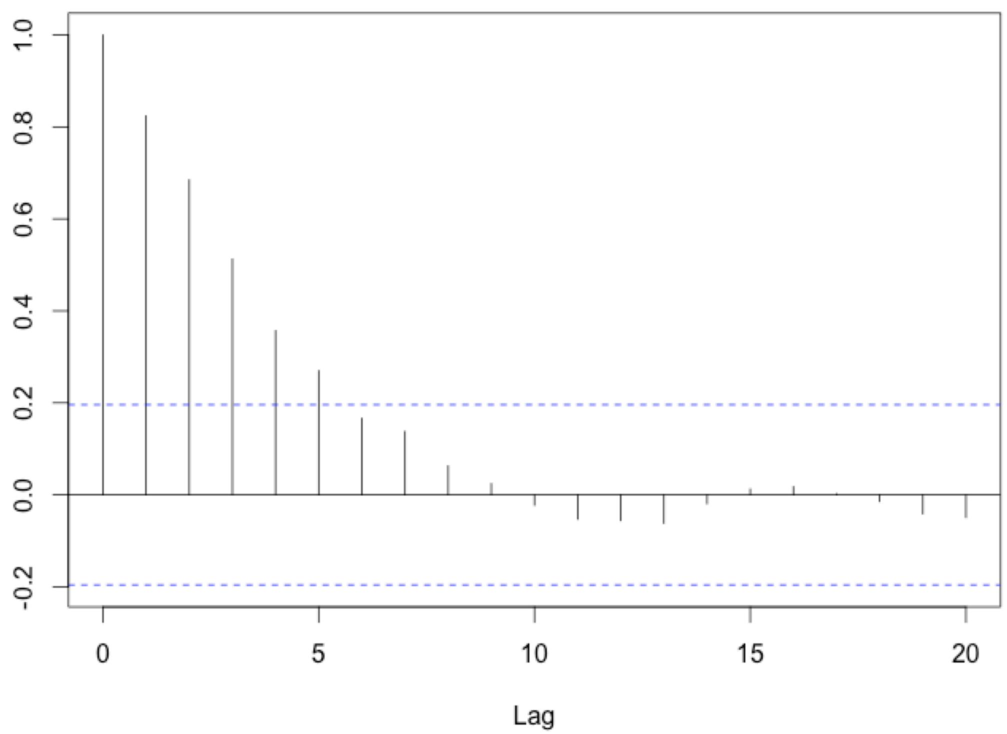
Autocorrelation plots

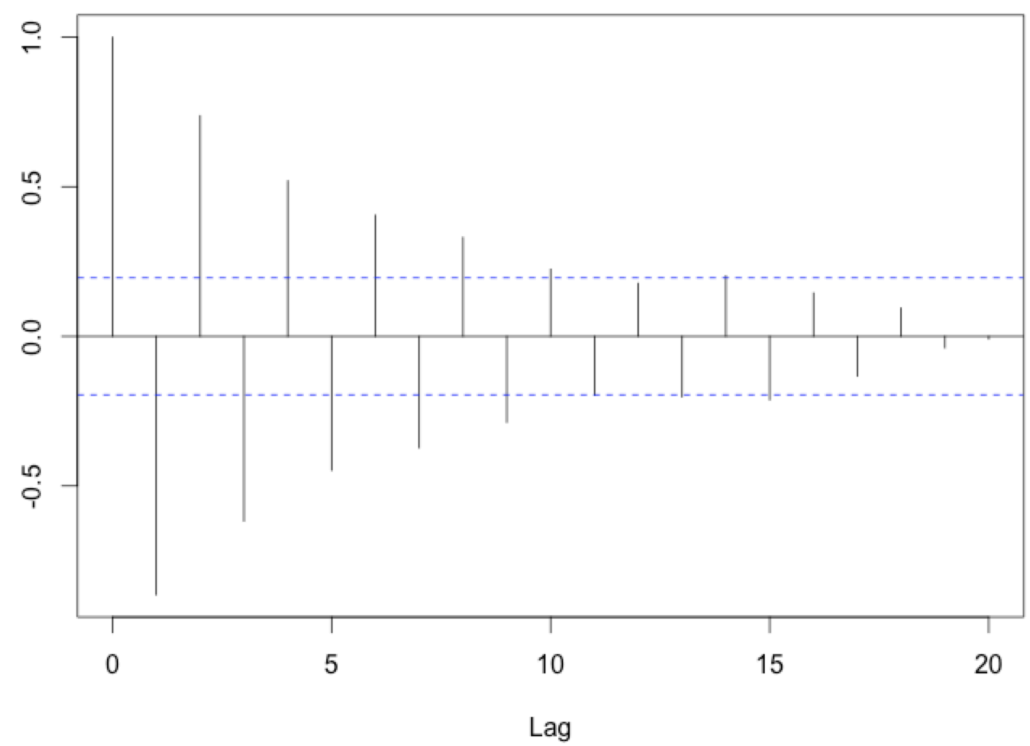
- A plotting method with which you can assess autocorrelation and moving averages

- Two plots
  - Autocorrelation
  - Partial autocorrelation

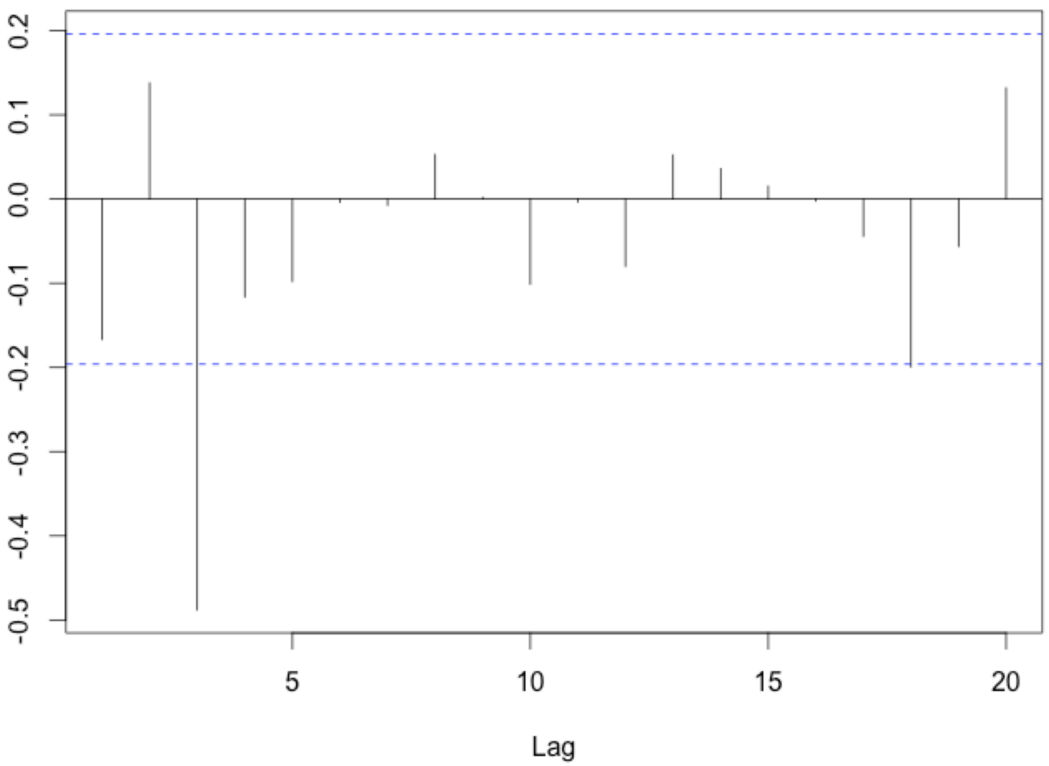
Model	ACF	Partial ACF
No autocorrelation	All zeros	All zeros
Autoregressive (p)	Exponential Decay	p significant lags before dropping to zero
Moving Average (q)	q significant lags before dropping to zero	Exponential Decay
Both (p,q)	Decay after q^{th} lag	Decay after p^{th} lag

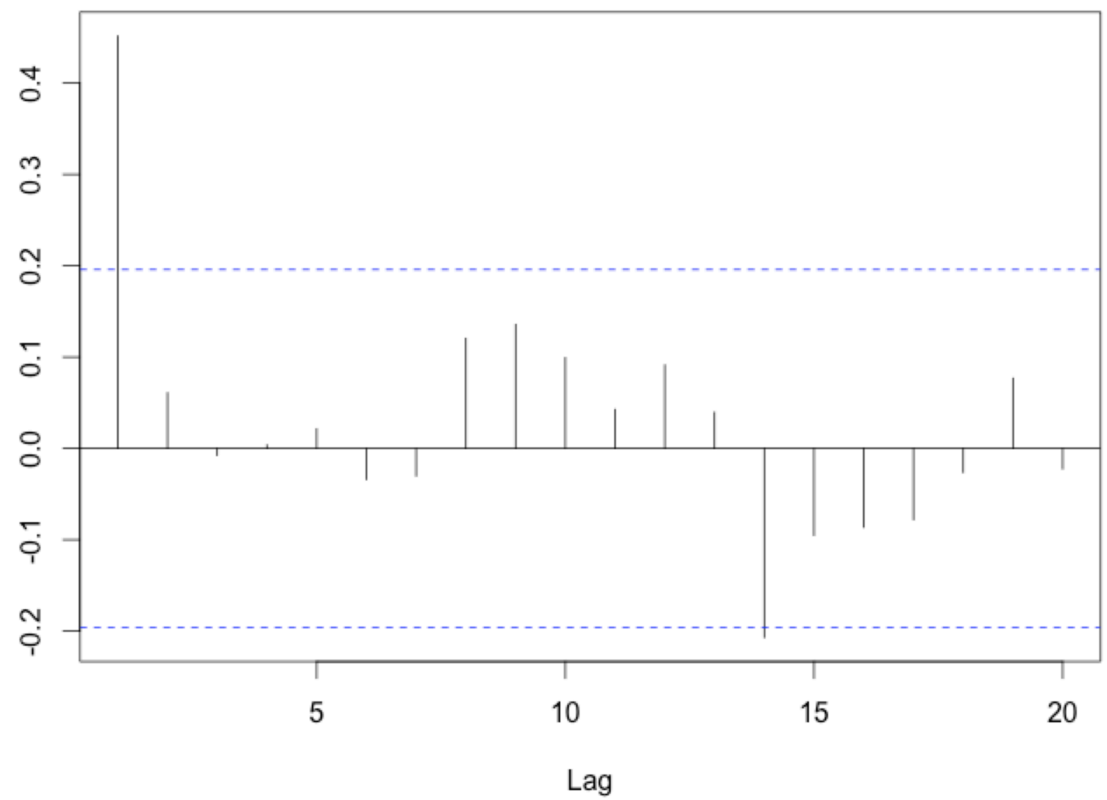
Examples of Exponential Decay

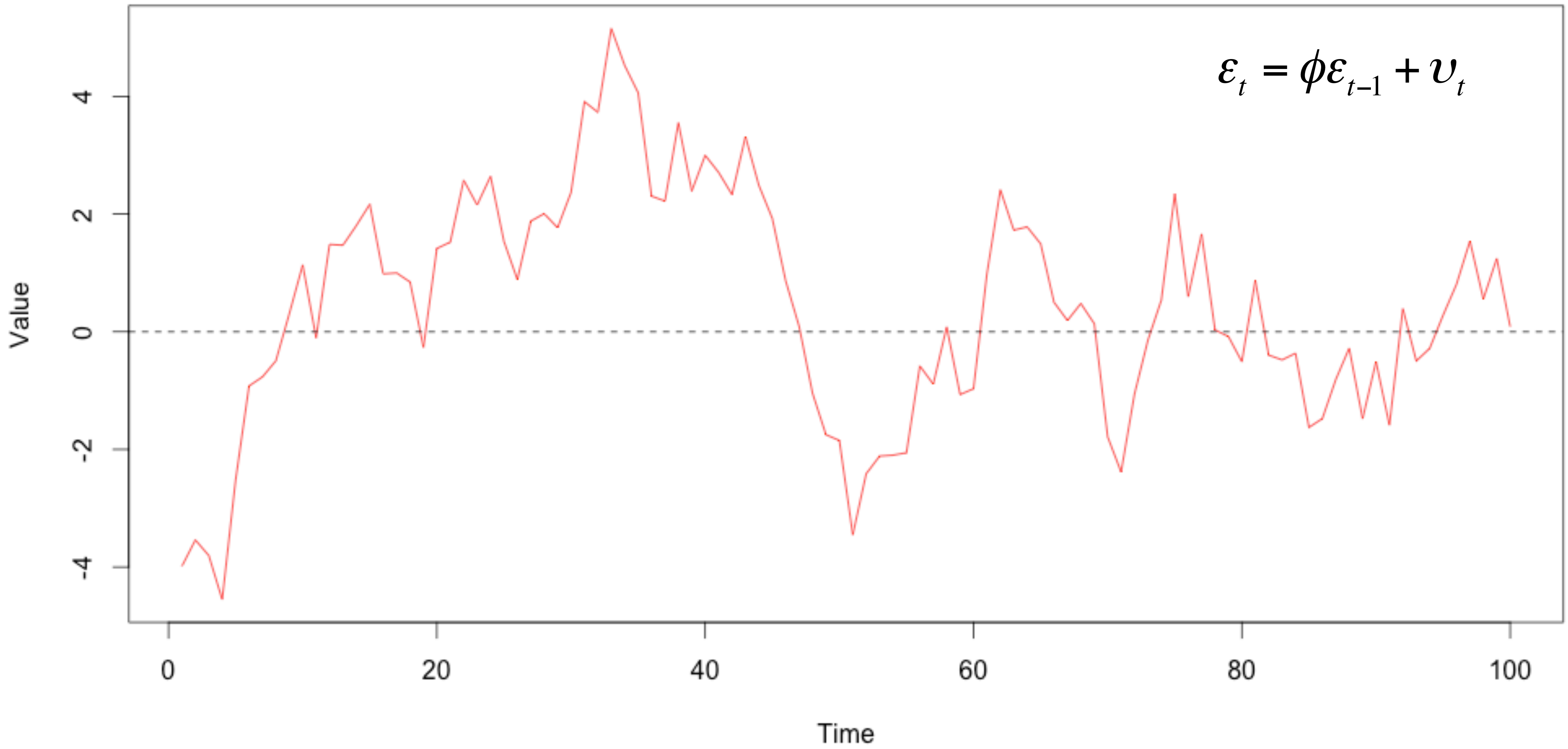


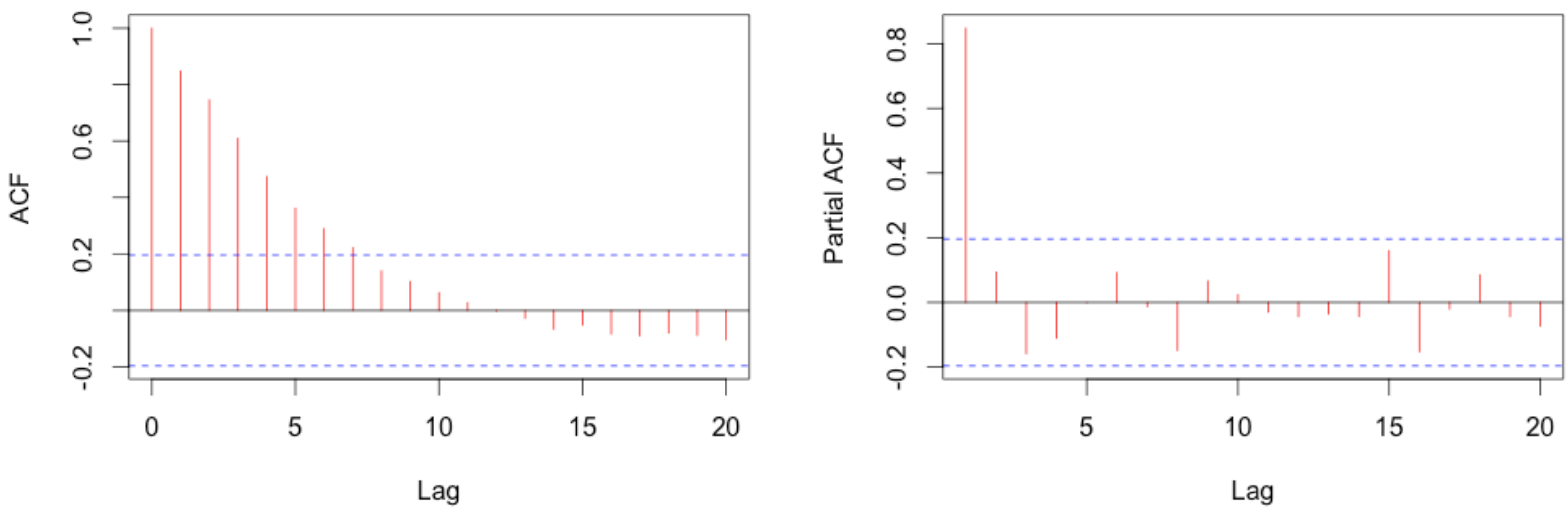


Examples of Significant Lags






$\varepsilon_t = \phi\varepsilon_{t-1} + \upsilon_t$
Value
Time
4
2
0
-2
-4
0
20
40
60
80
100

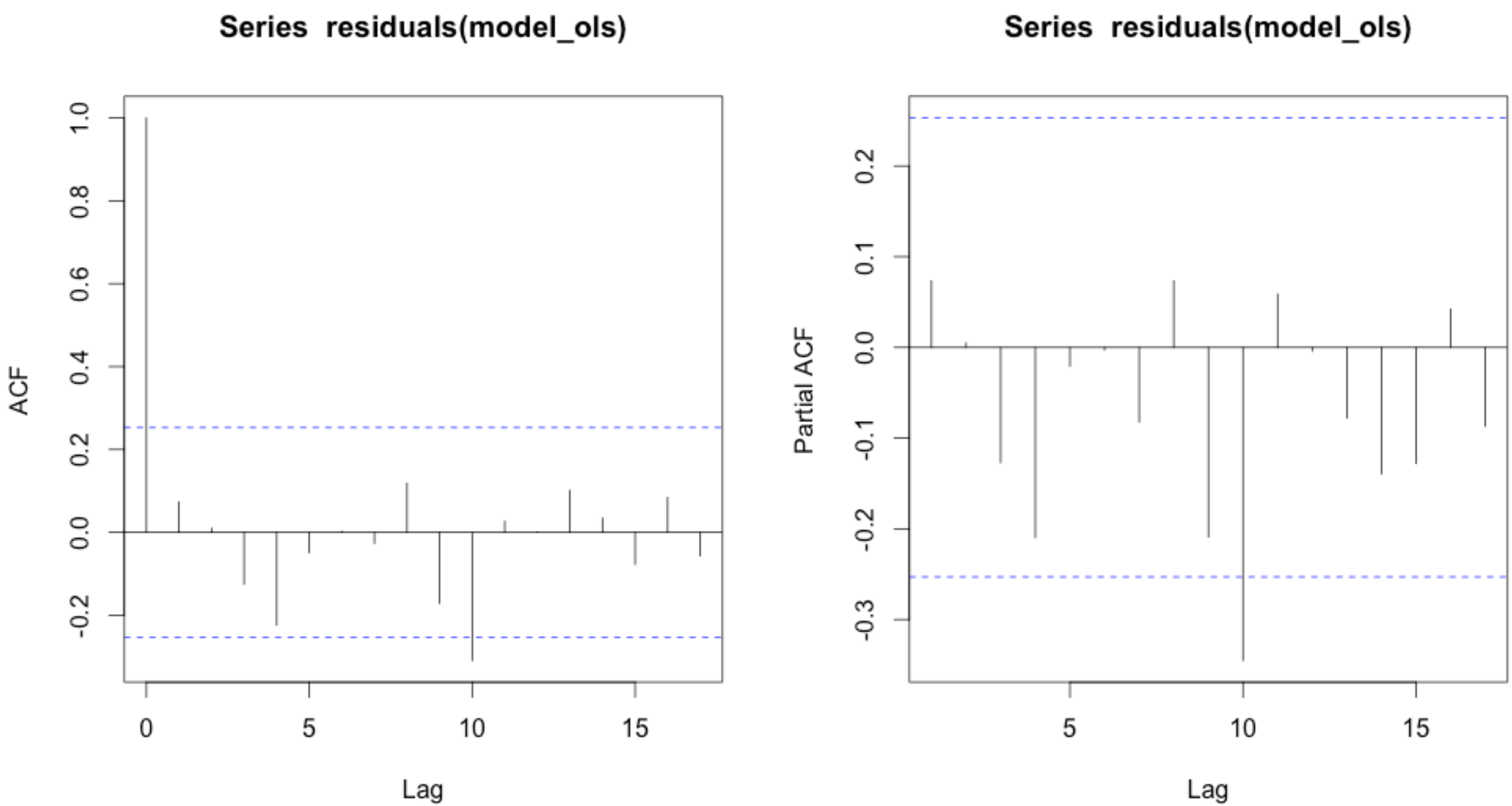


ACF
1.0
0.6
0.2
-0.2
0
5
10
15
20
Lag
Partial ACF
0.8
0.6
0.4
0.2
0.0
-0.2
5
10
15
20
Lag

Producing ACF plots

- Use the following code on our nile data:

```
# Set plotting to two records on one page
par(mfrow=c(2,1))

# Produce plots
acf(residuals(ols))
acf(residuals(ols),type='partial')
```


Series residuals(model_ols)
ACF
Lag
Series residuals(model_ols)
Partial ACF
Lag

STEP 8: RUN THE FINAL MODEL

Running the final model

- Generalized Least Squares fits a linear model with AR() and MA() processes
- Function *gls()* uses similar specification to *lm( )*

```
# Fit the GLS regression model
model_p10 <- gls(flow ~ time + level + trend,
   data=data,
   correlation=corARMA(p=10,form=~time),
   method="ML")

summary(model_p10)
```

```
Coefficients:
                Value     p-value
(Intercept) 3348.033     0.0000
time           6.745     0.2871
level       -738.245     0.0000
trend        -14.210     0.0397
```

Interpretation: after the weather change, there was a sustained drop in average monthly water flow of 738 million cubic meters. There was a small drop in the trend of 14.2 million cubic meters per month.

Model checking

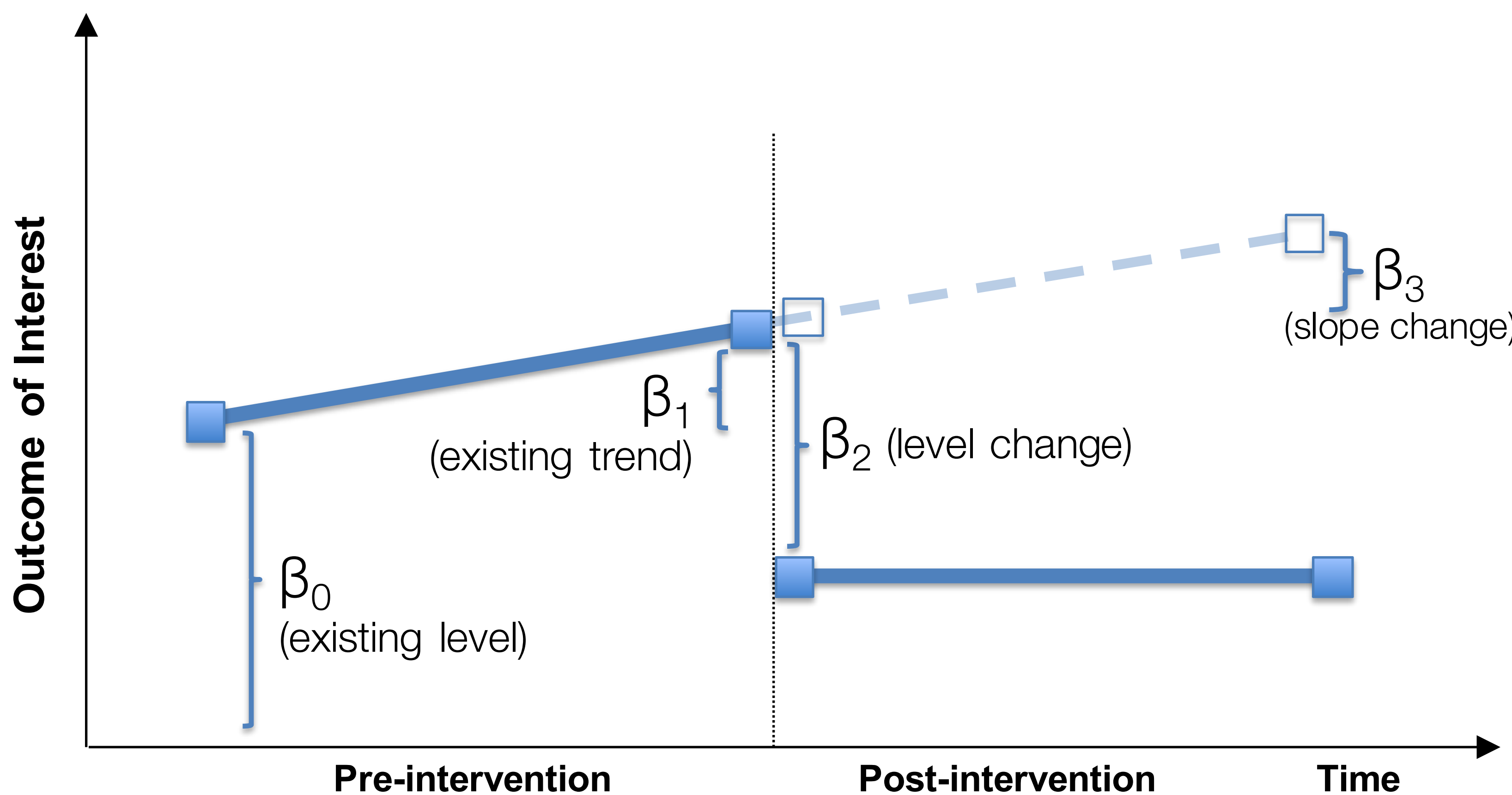
- To check the specification of the correlation structure, you can formally test the inclusion of further autoregressive parameters

```
# Likelihood-ratio  tests to check AR process

model_p11 <- update(model_p10,correlation=corARMA(p=11,form=~time))
anova(model_p10,model_p11)
```

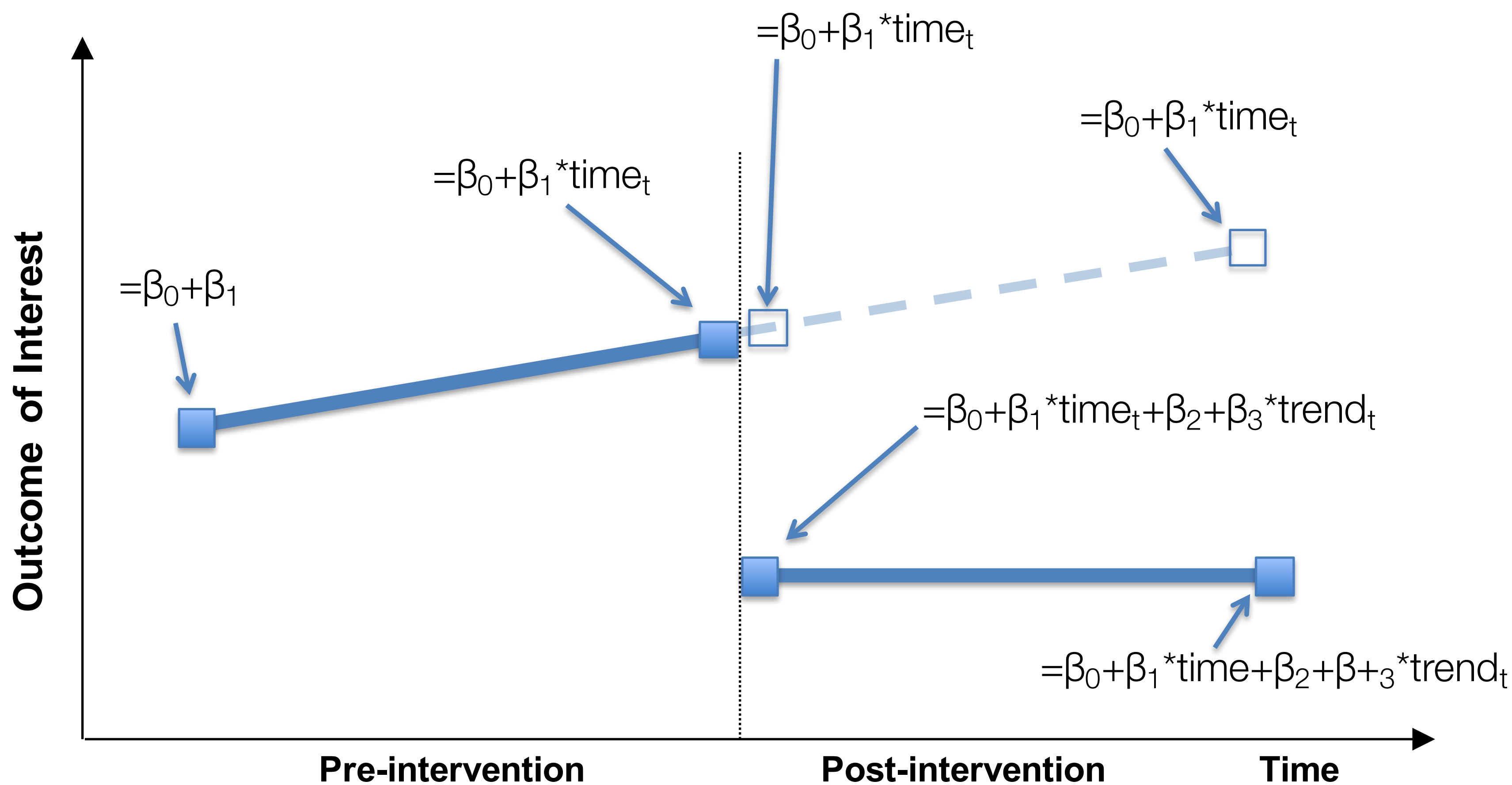
```
> model_p11 <- update(model_p10,correlation=corARMA(p=11,form=~time))
> anova(model_p10,model_p11)
          Model df      AIC      BIC    logLik   Test   L.Ratio p-value
model_p10     1 15 913.4212 944.8364 -441.7106
model_p11     2 16 915.0514 948.5609 -441.5257 1 vs 2 0.3698229  0.5431

> model_p10q1 <- update(model_p10,correlation=corARMA(p=10,q=1,form=~time))
> anova(model_p10,model_p10q1)
            Model df      AIC      BIC    logLik   Test   L.Ratio p-value
model_p10       1 15 913.4212 944.8364 -441.7106
model_p10q1     2 16 915.1611 948.6706 -441.5805 1 vs 2 0.2601591    0.61
```

STEP 9: PLOT THE RESULTS

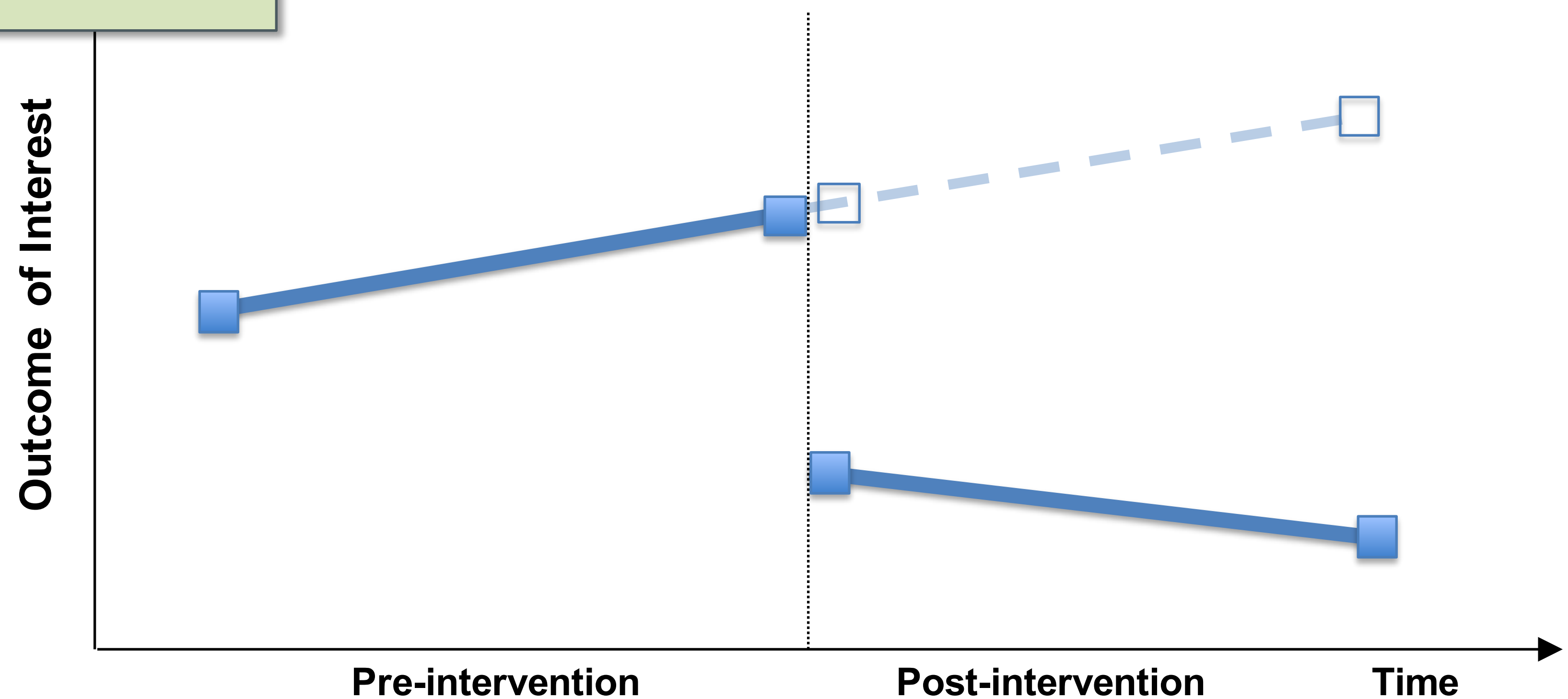
$$outcome_{jt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot level_j + \beta_3 \cdot trend_{jt} + \varepsilon_{jt}$$



$$outcome_{jt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot level_j + \beta_3 \cdot trend_{jt} + \varepsilon_{jt}$$



```
Coefficients:
                  Value
(Intercept)  3348.033
time            6.745
level        -738.245
trend         -14.210
```



Plot the results

```
# Produce the plot, first
plotting the raw data points

plot(data$time,data$flow,
       ylim=c(0,4500),
       ylab=”Water Flow",
       xlab="Year",
       pch=20,
       col=”pink",
       xaxt="n”)
```

Plot continued

```
# Plot dates on x-axis
axis(1, at=1:60, labels=data$year)

# Add line indicating weather pattern change
abline(v=27.5,lty="dotted")
```

Completing the plot

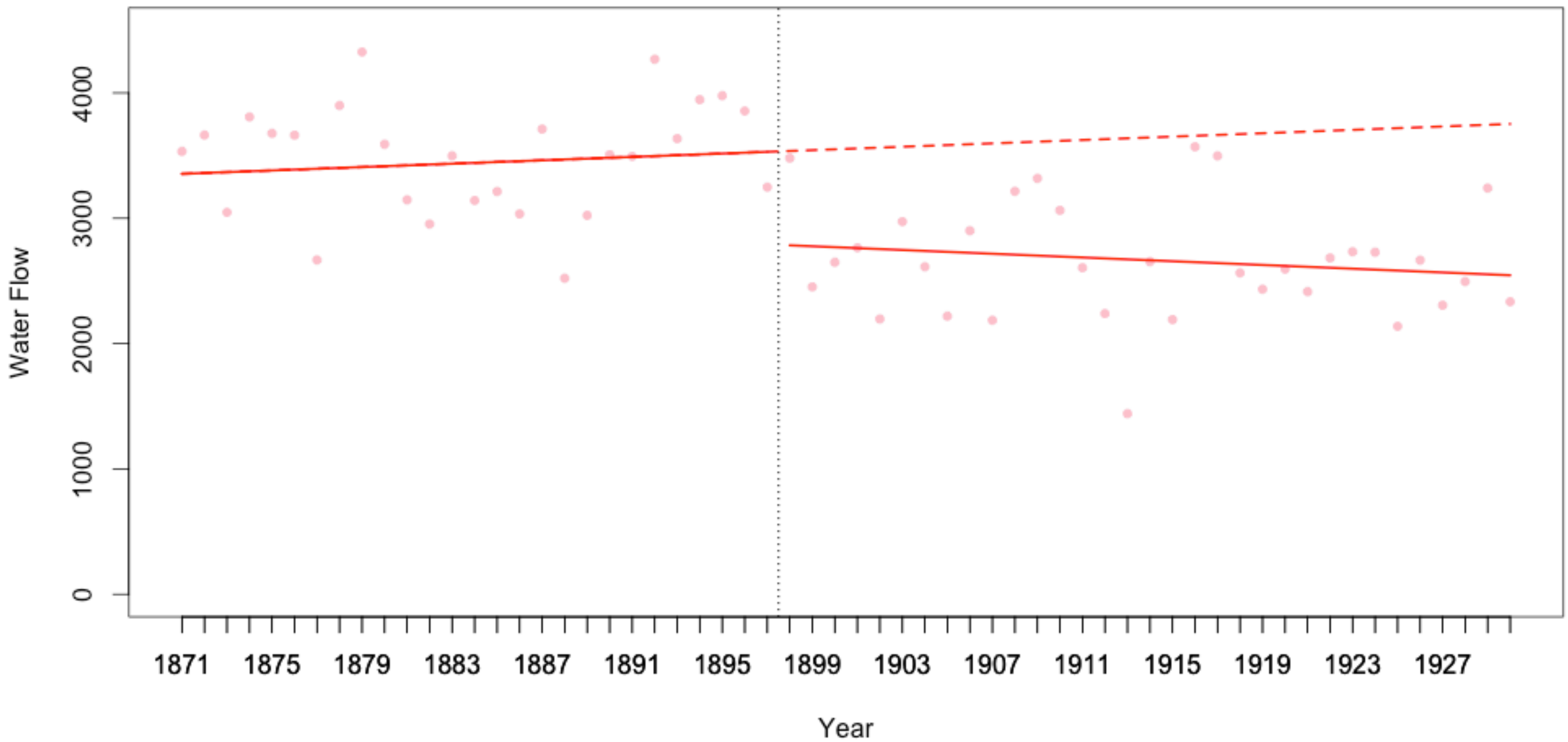
- For observed, use *fitted ( )*

```
# Plot the first line segment
lines(data$time[1:27], fitted(model_p10)[1:27], col="red",lwd=2)

# Plot the second line segment
lines(data$time[28:60], fitted(model_p10)[28:60], col="red",lwd=2)
```

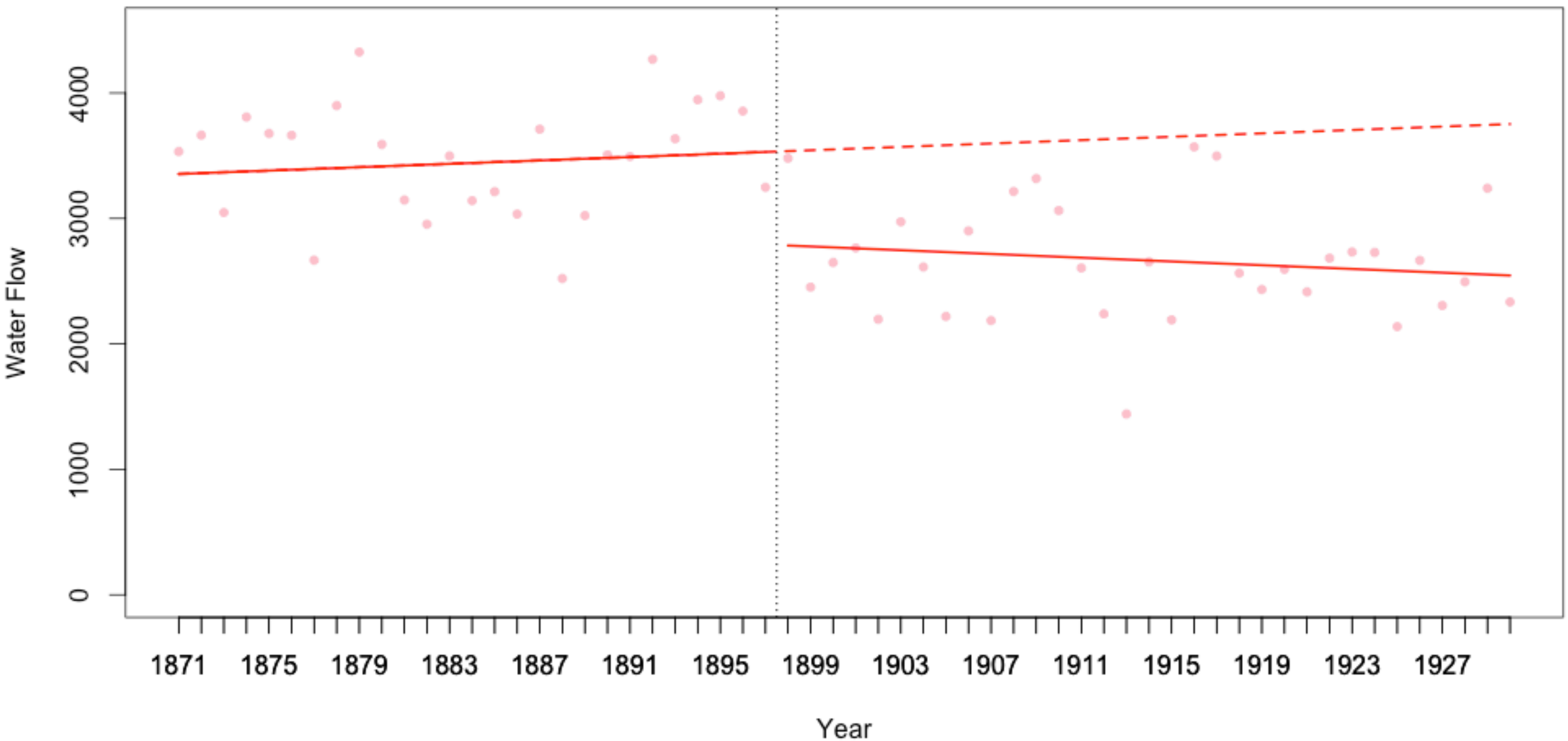
Plotting the counterfactual

- For counterfactual, use *segments*(x_1,y_1,x_2,y_2)

```
segments(1,
         model_p10$coef[1]+model_p10$coef[2],
         60,
         model_p10$coef[1]+model_p10$coef[2]*60,
         lty=2,
         lwd=2,
         col='red')
```

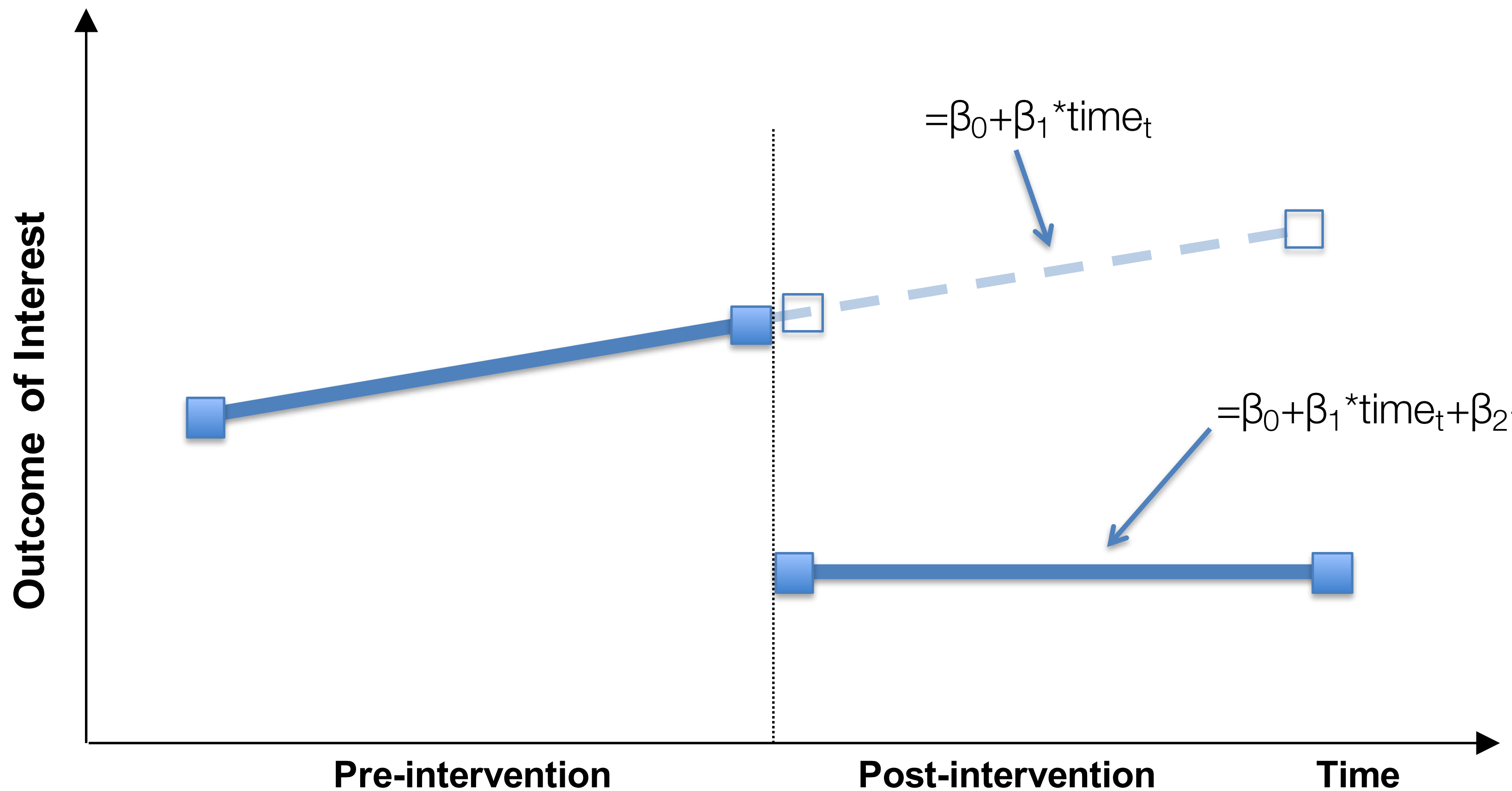

Water Flow
0
1000
2000
3000
4000
1871
1875
1879
1883
1887
1891
1895
1899
1903
1907
1911
1915
1919
1923
1927
Year

STEP 10: PREDICTED CHANGES



Water Flow
0
1000
2000
3000
4000
1871
1875
1879
1883
1887
1891
1895
1899
1903
1907
1911
1915
1919
1923
1927
Year

$$outcome_{jt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot level_j + \beta_3 \cdot trend_{jt} + \varepsilon_{jt}$$



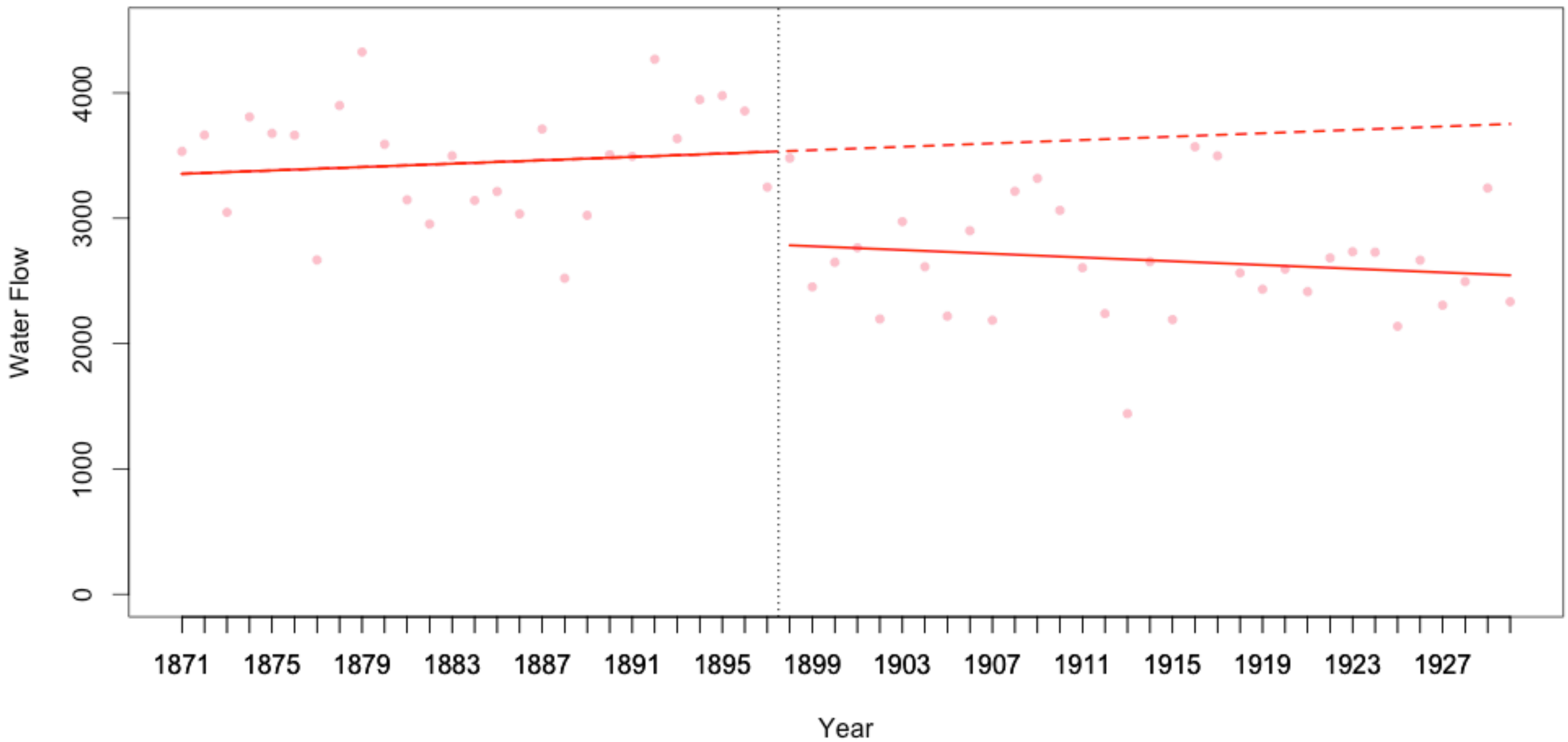
Step 10: Predicted changes

- Using the model coefficients, you can predict absolute and relative changes

```
# Predicted value at 25 years after the weather change
pred <- fitted(model_p10)[52]

# Then estimate the counterfactual at the same time
point
cfac <- model_p10$coef[1] + model_p10$coef[2]*52

# Absolute change at 25 years
pred - cfac

# Relative change at 25 years
(pred - cfac) / cfac
```

```
> # Absolute change at 25 years
> pred - cfac
       52
-1093.483

> # Relative change at 25 years
> (pred - cfac) / cfac
        52
-0.2956351
```

Interpretation: In the 25th year after the weather change, the average monthly water flow was 1093 million cubic meters less than would have been expected if the weather had not changed. This represented a 29.6% reduction.



Water Flow
0
1000
2000
3000
4000
1871
1875
1879
1883
1887
1891
1895
1899
1903
1907
1911
1915
1919
1923
1927
Year