ITSx: Policy Analysis Using Interrupted Time Series

Week 3 Slides

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

Layout of the weeks

1. Introduction, setup, data sources
2. 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

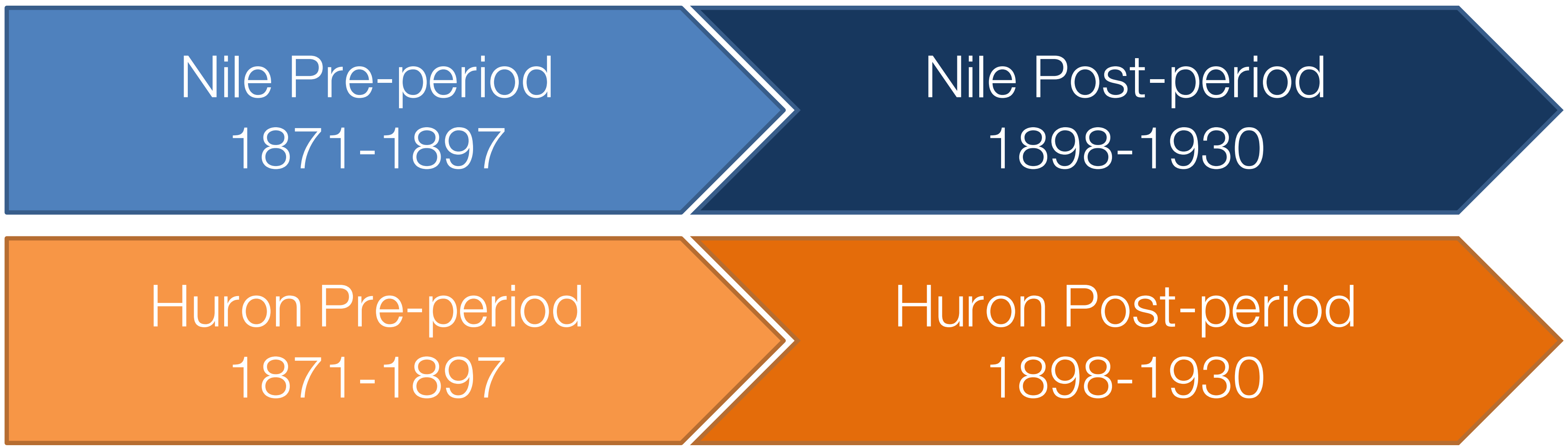
Source: http://commons.wikimedia.org/wiki/File:River_Nile_map.svg


Canada
USA

Research Question

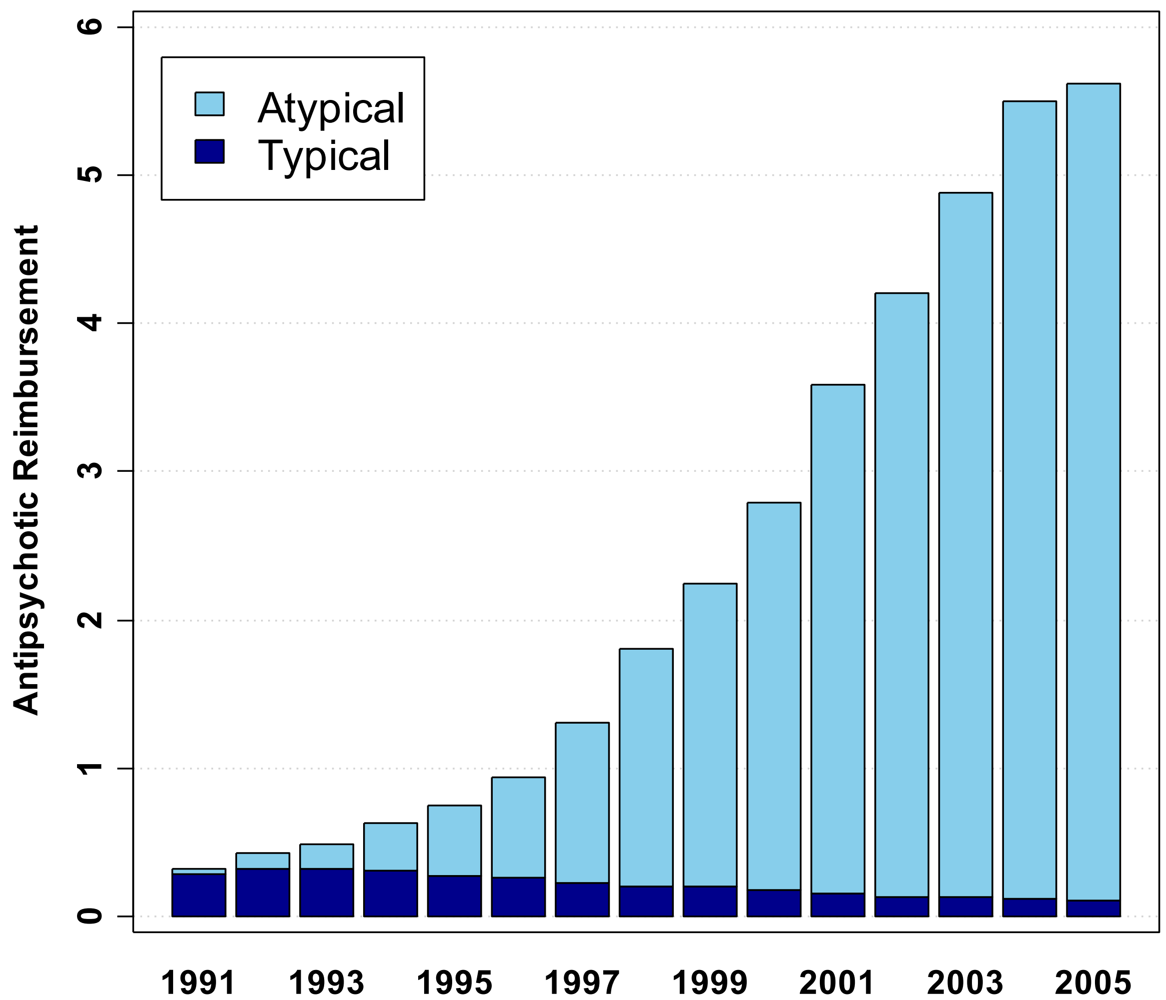
- Previous work has suggested weather patterns shifted in 1898 (Cobb 1978)

- What was the impact of weather changes on annual water flow levels in the Nile?


Nile Pre-period
1871-1897
Nile Post-period
1898-1930
Huron Pre-period
1871-1897
Huron 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 Department of Health & Human Resources
BUREAU FOR MEDICAL SERVICES

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)

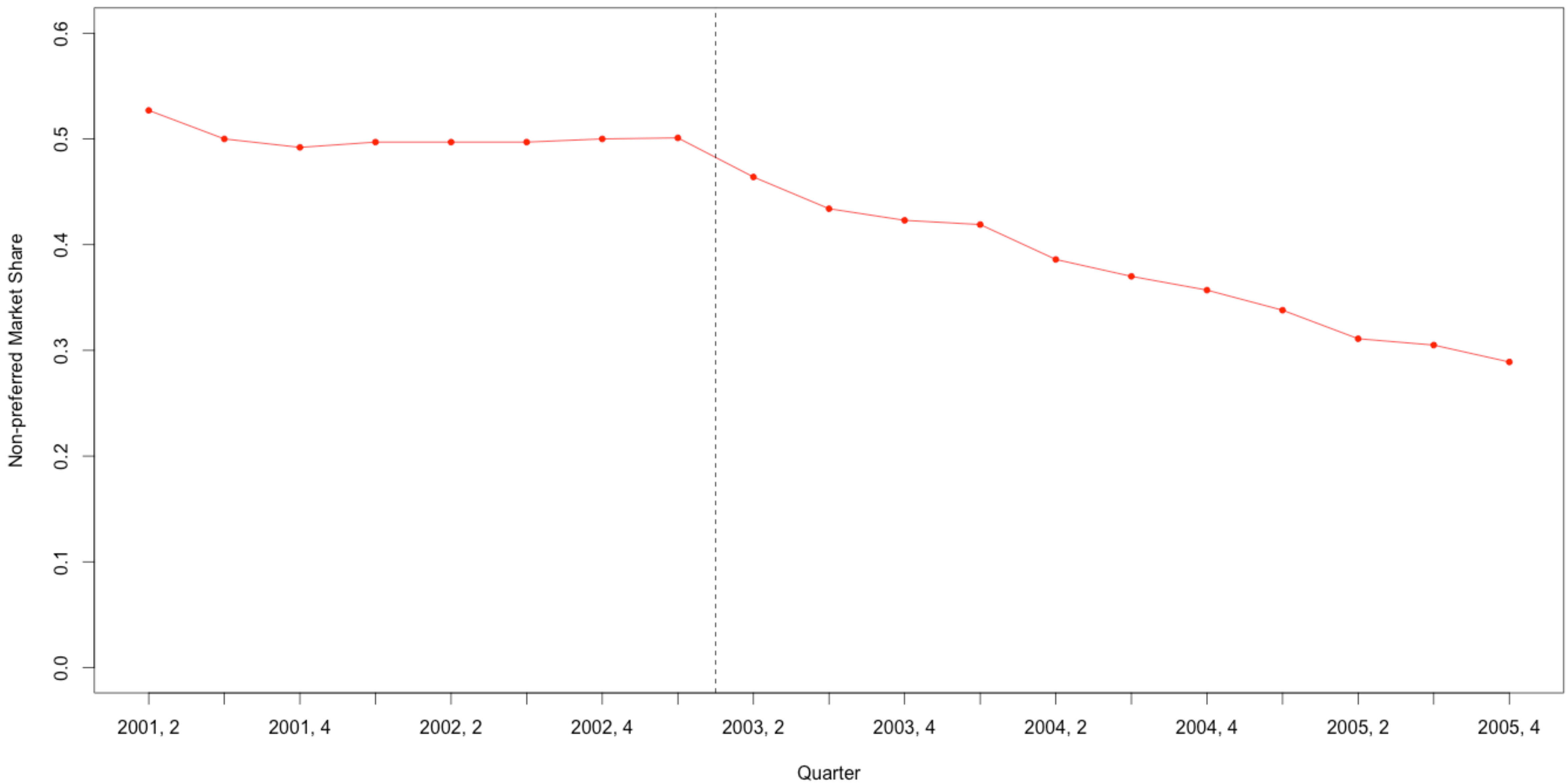
Prescriber Address (Street)	(City)	(State)	(Zip)
		West Virginia	

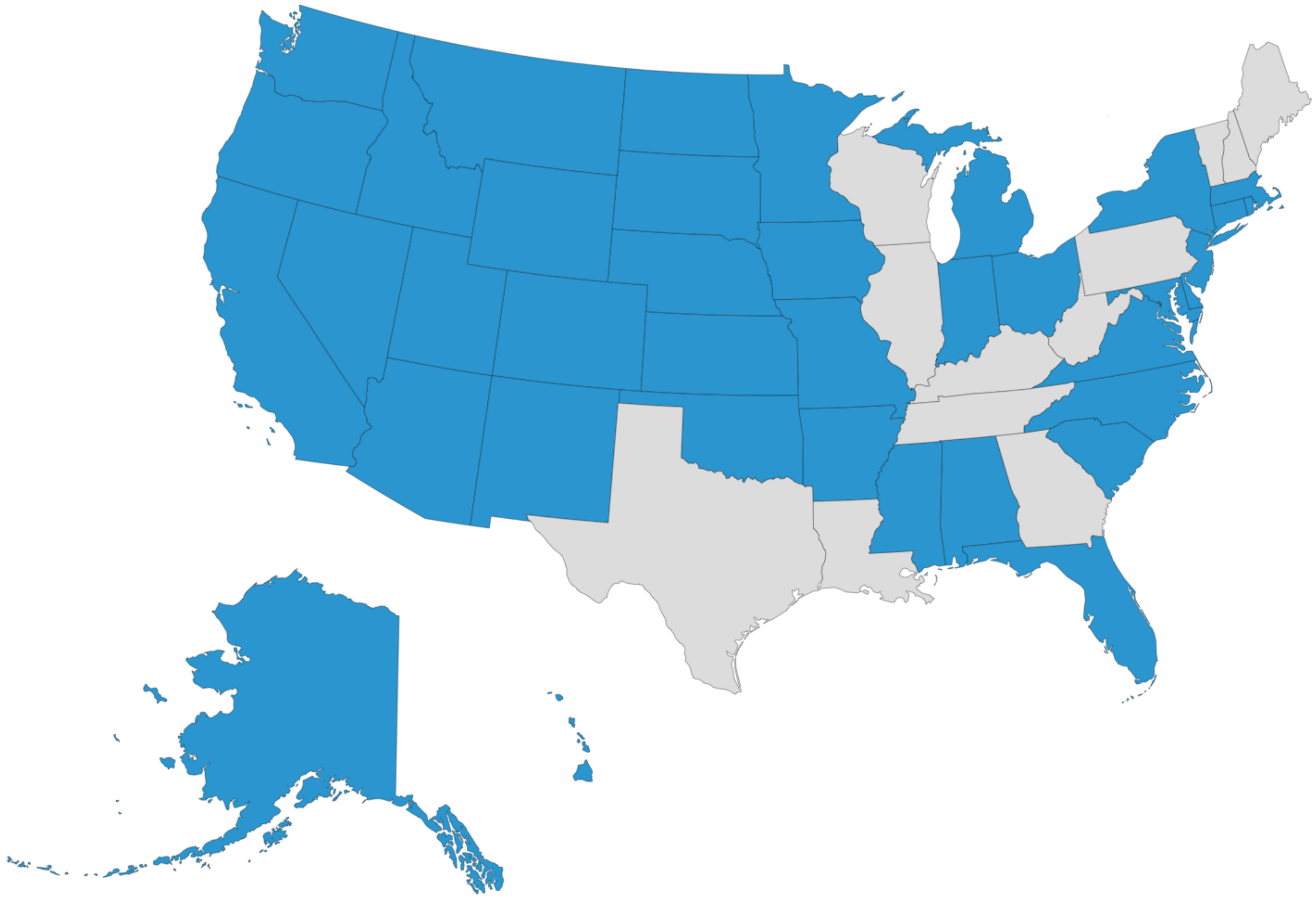
Prescriber 10-Digit NPI#	Phone # (111-222-3333)	Fax # (111-222-3333)

Pharmacy Name (if applicable)

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

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


Non-preferred Market Share
0.0
0.1
0.2
0.3
0.4
0.5
0.6
2001, 2
2001, 4
2002, 2
2002, 4
2003, 2
2003, 4
2004, 2
2004, 4
2005, 2
2005, 4
Quarter



STEP 4: SETUP DATA

Data Setup

- One data row for each group, in each time period

- Variables:

Existing Trend	**Existing Level Difference**
Post-intervention Level Change	**Existing Trend Difference**
Post-intervention Trend Change	**Level Change Difference**
Outcome of interest	**Trend Change Difference**

year	nile	flow
1871	1	3533.9
1872	1	3664.3
…	…	…
1896	1	3856.0
1897	1	3248.2
1898	1	3479.8
1899	1	2453.0
…	…	…
1929	1	3241.1
1930	1	2334.4
1871	0	2120.0
1872	0	2273.3

year	nile	flow	time	level	trend	niletime	nilelevel	niletrend
1871	1	3533.9	1	0	0	1	0	0
1872	1	3664.3	2	0	0	2	0	0
…	…	…	…	…	…	…	…	…
1929	1	3241.1	59	1	32	59	1	32
1930	1	2334.4	60	1	33	60	1	33
1871	0	2120.0	1	0	0	0	0	0
1872	0	2273.3	2	0	0	0	0	0
…	…	…	…	…	…	…	…	…
1929	0	2060.8	59	1	32	0	0	0
1930	0	1796.7	60	1	33	0	0	0

STEP 5: VISUALLY INSPECT DATA

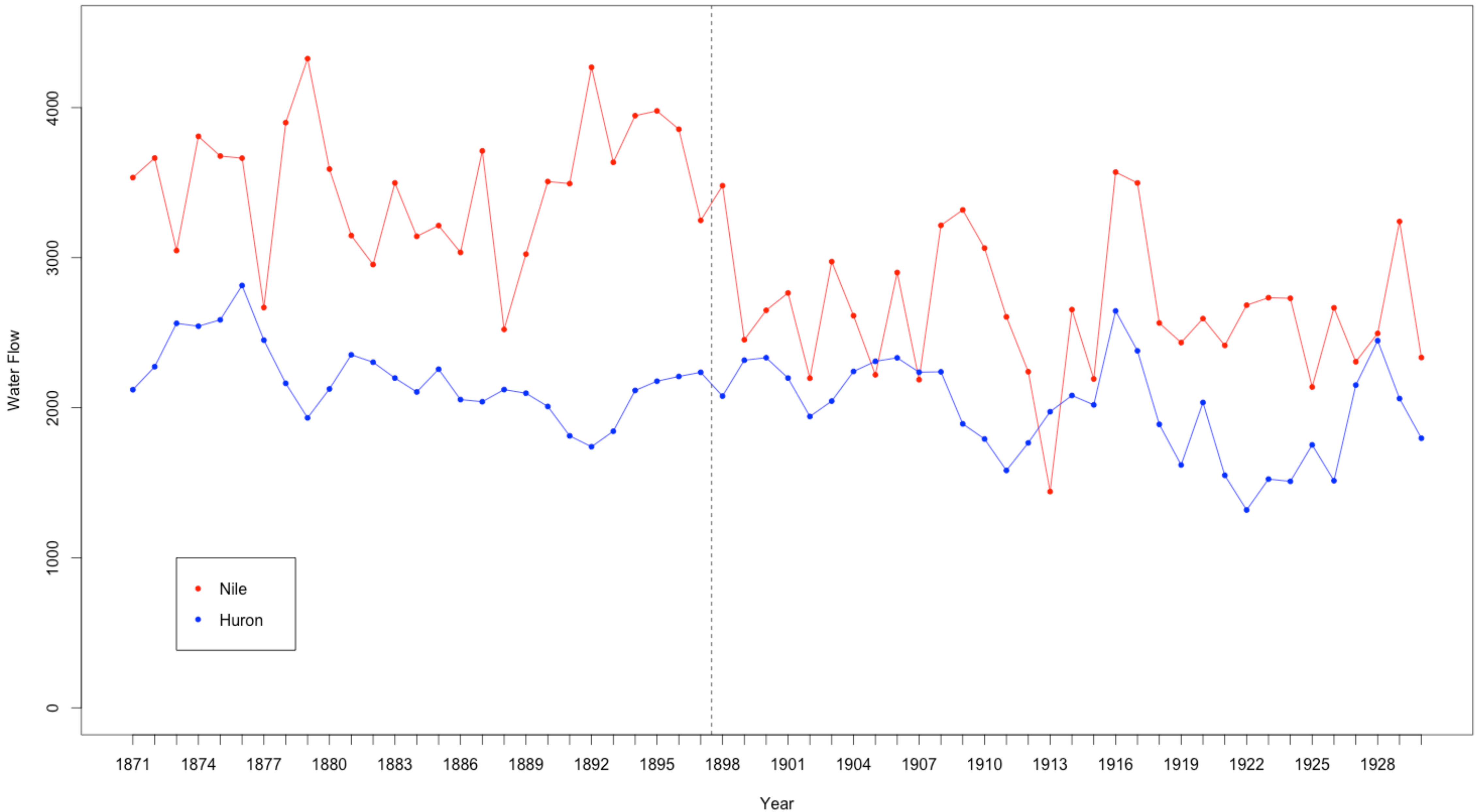
Visually Inspect Data

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

- Add to this: suitability of the control group

```
 # Plot the time series for the Nile river at Aswan
plot(data$time[1:60],data$flow[1:60],
     ylab="Water Flow",
     ylim=c(0,4500),
     xlab="Year",
     type="l",
     col="red",
     xaxt="n")

# Add in control group flow into Lake Huron
points(data$time[61:120],data$flow[61:120],
       type='l',
       col="blue")

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

```
# Add in the points for the figure
points(data$time[1:60],data$flow[1:60],
       col="red",
       pch=20)

points(data$time[61:120],data$flow[61:120],
       col="blue",
       pch=20)

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

# Add in a legend
legend(x=3, y=1000, legend=c("Nile","Huron"),
   col=c(”red",”blue"),pch=20)
```



Water Flow
0
1000
2000
3000
4000
Nile
Huron
1871
1874
1877
1880
1883
1886
1889
1892
1895
1898
1901
1904
1907
1910
1913
1916
1919
1922
1925
1928
Year

STEP 6: PRELIMINARY ANALYSIS

Preliminary analysis

- As with last week, start with 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*, for group *k*, at time *t*:

$$outcome_{jkt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot group_k +$$
$$\beta_3 \cdot group_k \cdot time_t + \beta_4 \cdot level_{jt} +$$
$$\beta_5 \cdot trend_{jt} + \beta_6 \cdot level_{jt} \cdot group_k +$$
$$\beta_7 \cdot trend_{jt} \cdot group_k + \varepsilon_{jkt}$$

OLS Regression in R

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

# See summary of model output
summary(model_ols)

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

OLS Model Results

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2442.191	148.540	16.441	< 2e-16	***
time	-17.731	9.272	-1.912	0.05838	.
nile	965.745	210.068	4.597	1.13e-05	***
niletime	23.149	13.112	1.765	0.08021	.
level	256.428	193.936	1.322	0.18878	
trend	4.022	11.534	0.349	0.72796	
nilelevel	-1062.702	274.267	-3.875	0.00018	***
niletrend	-14.983	16.311	-0.919	0.36029	

STEP 7: AUTOCORRELATION

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

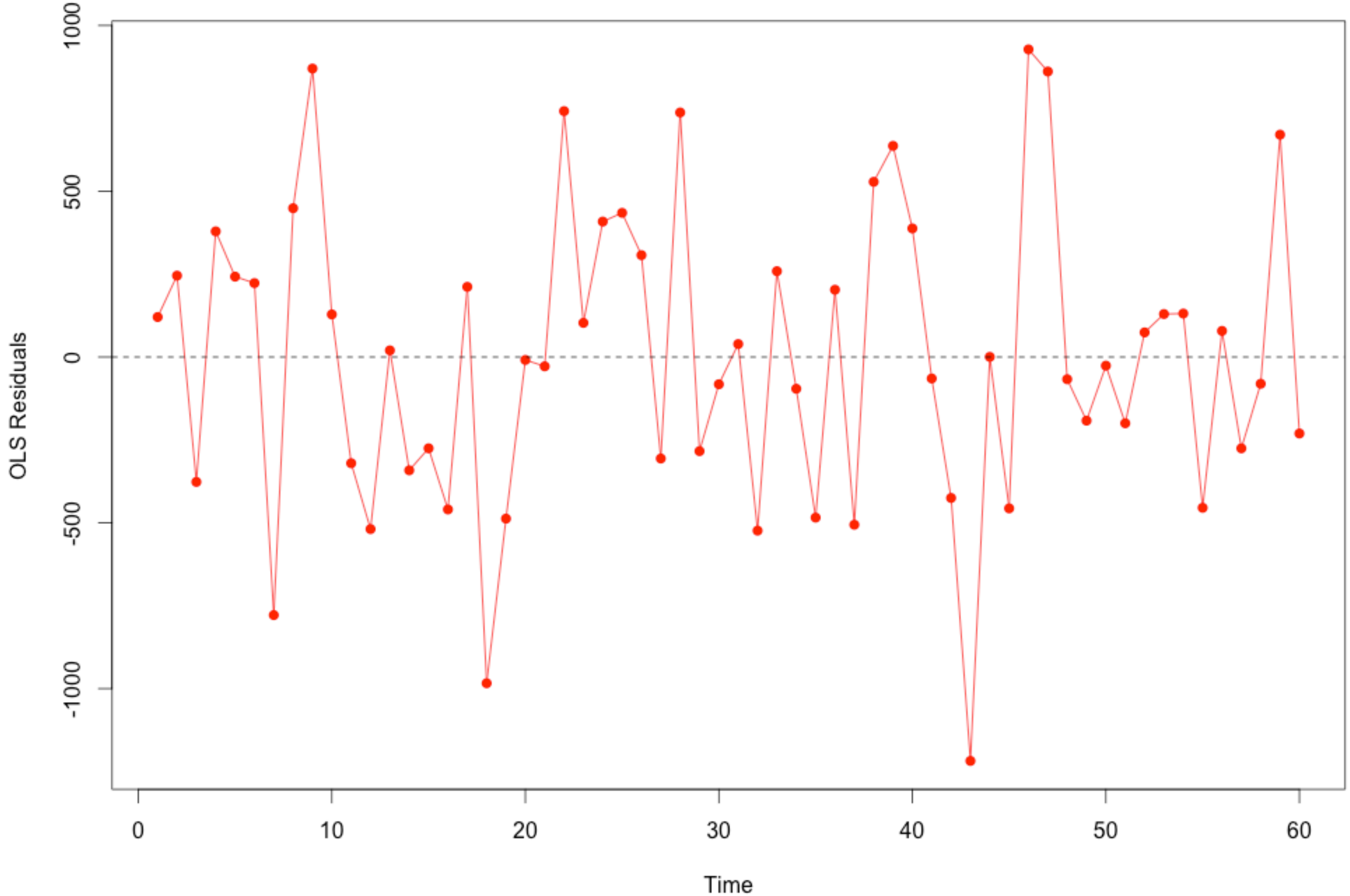
```
# Durbin-watson test, 12 time periods
dwt (model_ols,max.lag=12, alternative="two.sided")
```

```
> dwt(model_ols,max.lag=12,alternative="two.sided")
 lag Autocorrelation D-W Statistic p-value
   1      0.18937367      1.620277   0.010
   2      0.01902871      1.952186   0.442
   3     -0.12878820      2.211857   0.402
   4     -0.21783205      2.373411   0.062
   5     -0.14356400      2.215094   0.292
   6     -0.10262651      2.129624   0.414
   7     -0.05243658      1.983483   0.864
   8      0.07835392      1.701832   0.188
   9     -0.13712761      2.065004   0.438
  10     -0.22540820      2.233116   0.092
  11      0.07906429      1.616586   0.178
  12      0.08853936      1.574862   0.130
 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

```
# Graph the residuals
plot(data$time[1:60],
    residuals(model_ols)[1:60],
    type='o',
    pch=16,
    xlab='Time',
    ylab='OLS Residuals',
    col="red")
abline(h=0,lty=2)
```



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

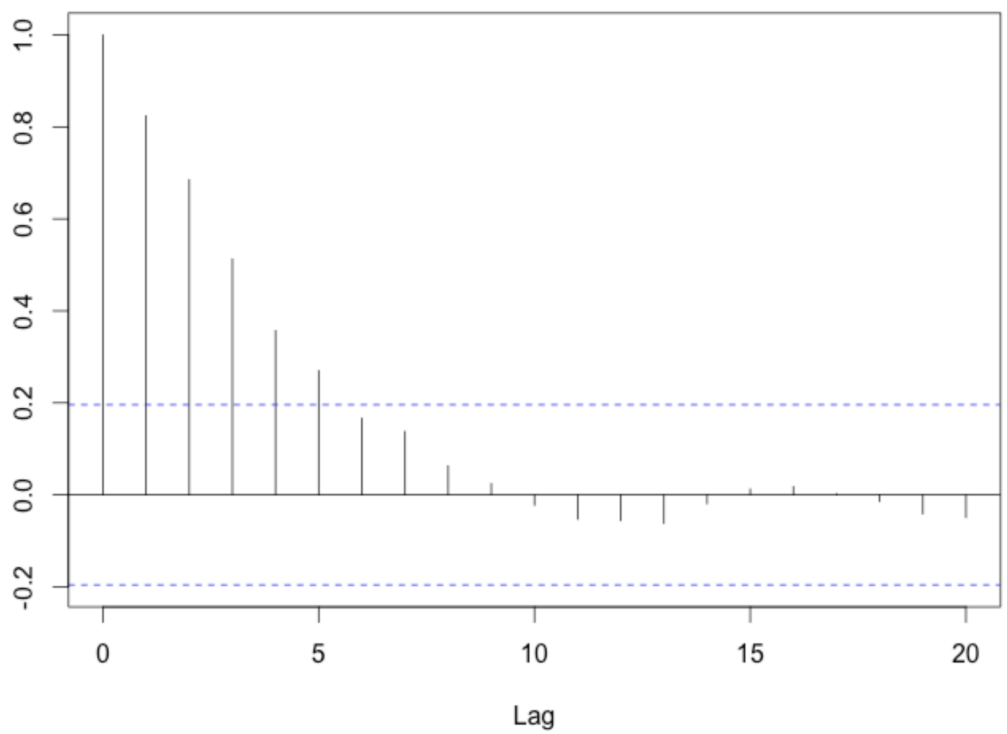
Autocorrelation plots

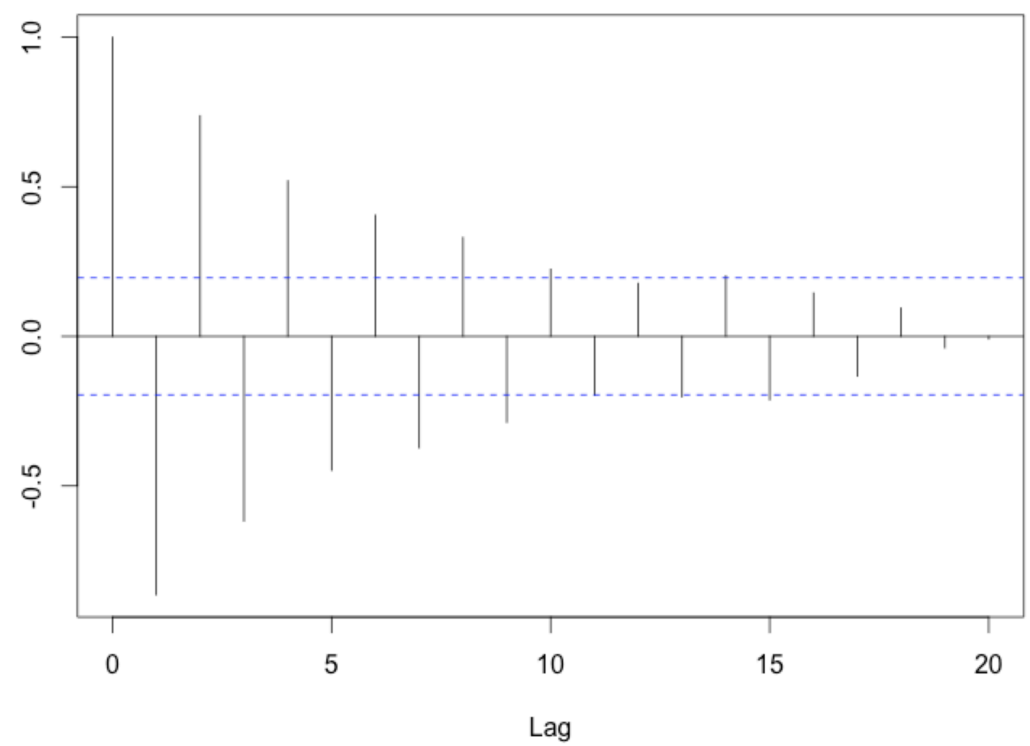
- A plotting method with which you can assess autocorrelation and moving averages

- Two plots
  - Autocorrelation
  - Partial autocorrelation

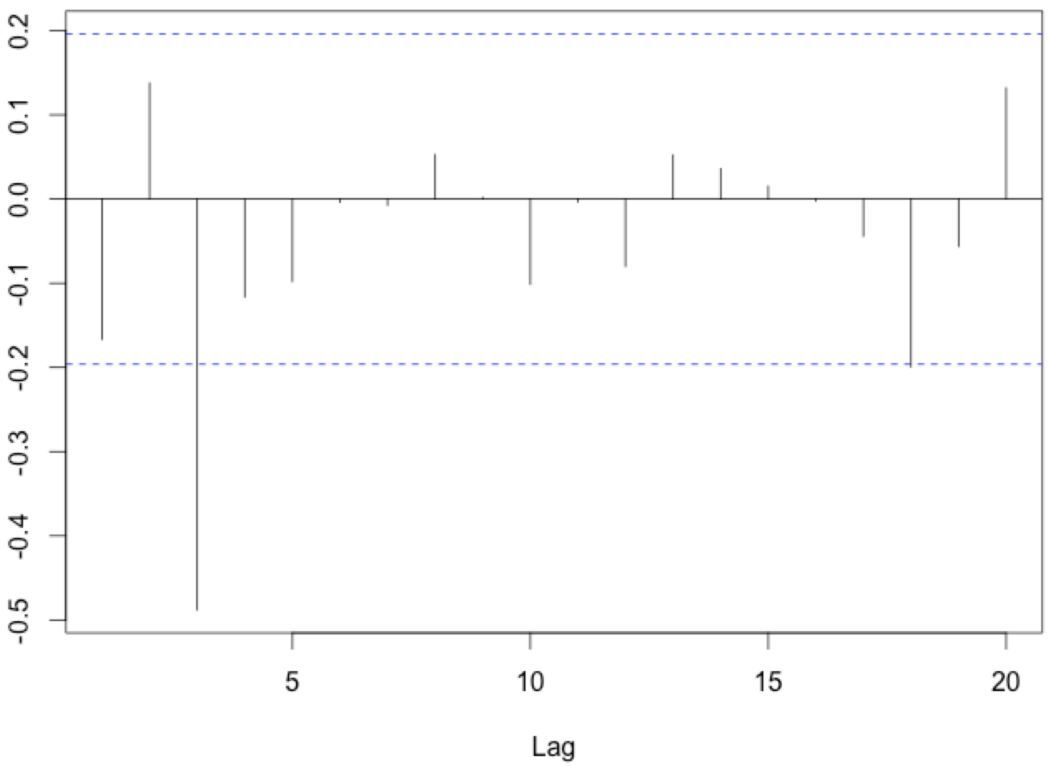
Model	ACF	Partial ACF
No autocorrelation	All zeros	All zeros
Autoregressive (p)	Exponential Decay	p significant lags before dropping to zero
Moving Average (q)	q significant lags before dropping to zero	Exponential Decay
Both (p,q)	Decay after q^{th} lag	Decay after p^{th} lag

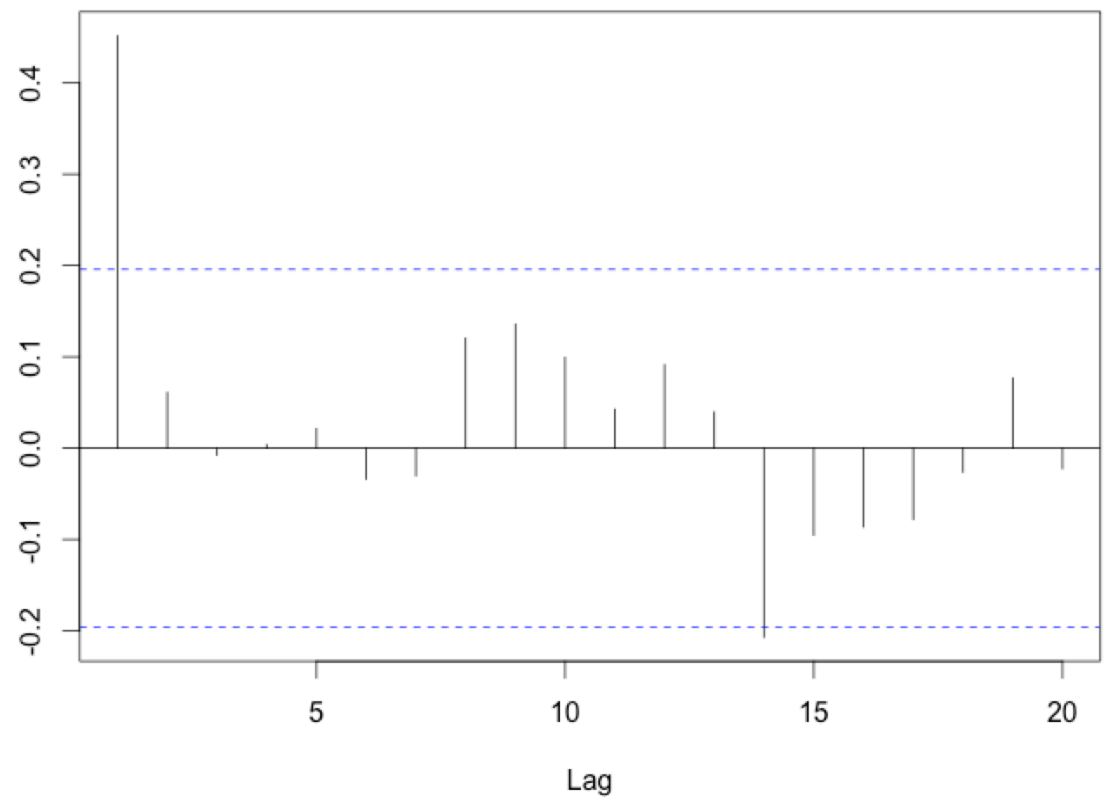
Examples of Exponential Decay





Examples of Significant Lags

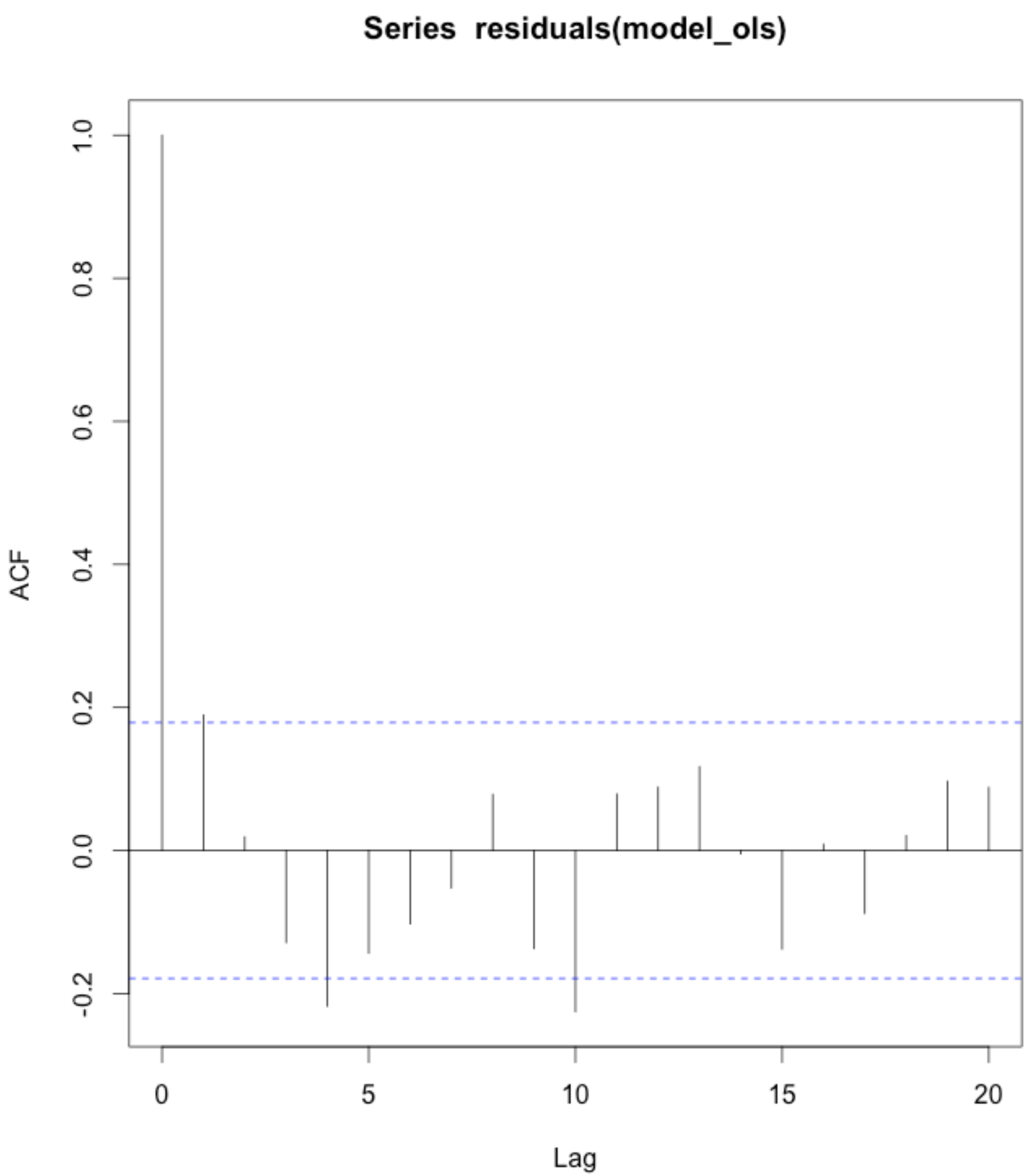


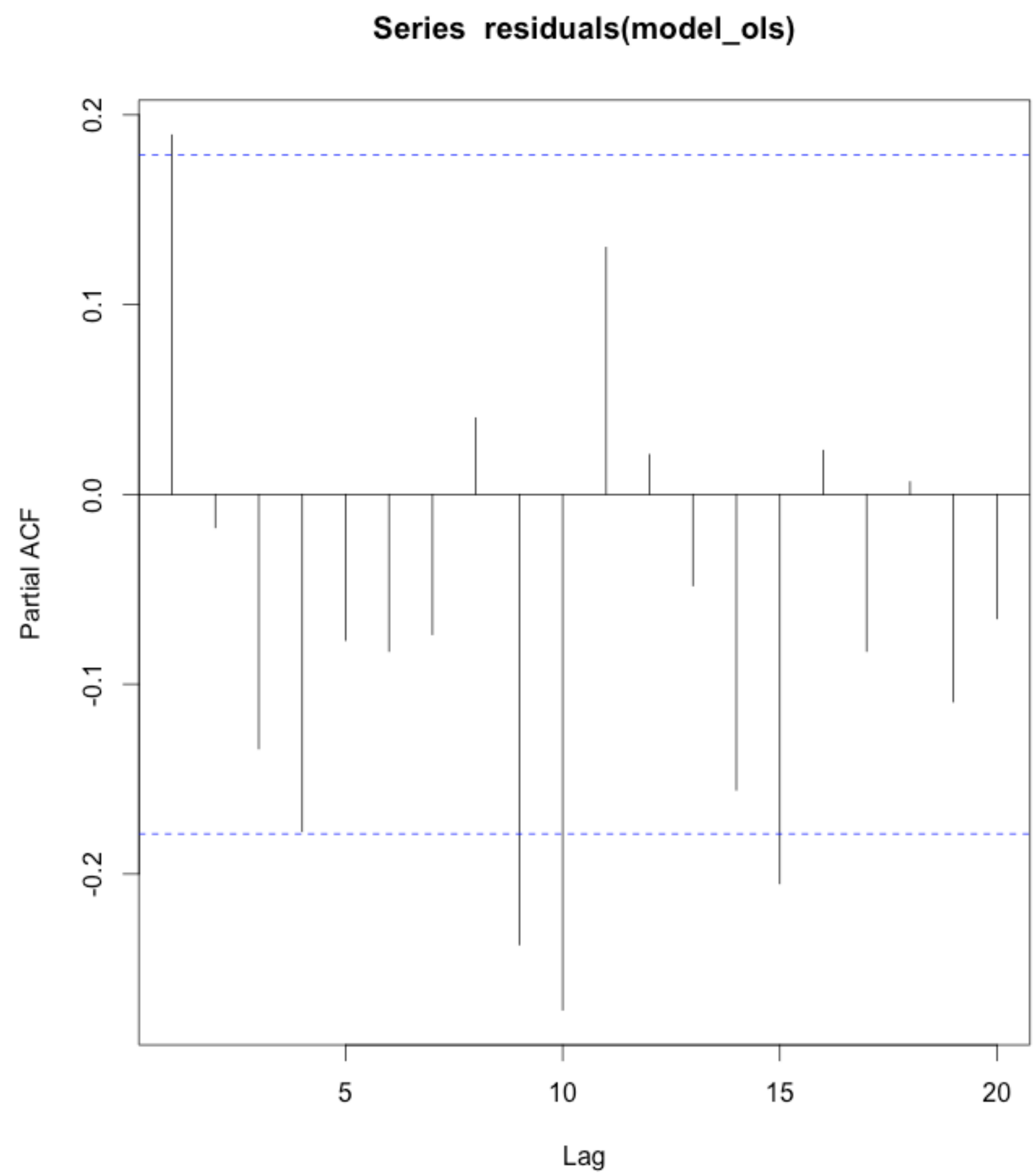


Producing ACF plots

- 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(model_ols))
acf(residuals(model_ols),type='partial')
```


Series residuals(model_ols)
ACF
Lag
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
20


Series residuals(model_ols)
Partial ACF
Lag
0.2
0.1
0.0
-0.1
-0.2
5
10
15
20

STEP 8: RUN THE FINAL MODEL

Running the final model

- As with last week, use *gls()*:

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

summary(model_p10)
confint(model_p10)
```

Coefficients:

	Value	Std.Error	t-value	p-value
(Intercept)	2464.7702	76.14223	32.37061	0.0000
time	-20.2181	4.85602	-4.16352	0.0001
nile	885.4280	107.68137	8.22267	0.0000
niletime	26.8242	6.86745	3.90600	0.0002
level	368.1143	101.97194	3.60996	0.0005
trend	2.1405	5.22455	0.40970	0.6828
nilelevel	-1101.1601	144.21010	-7.63580	0.0000
niletrend	-16.4864	7.38863	-2.23132	0.0277

Interpretation: after the weather change, there was a sustained drop in average monthly water flow of 1101 million cubic meters relative to the change in the St. Mary's river. There was also a small relative drop in the trend of 16.5 million cubic meters per month afterward.

Model checking

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

```
# Likelihood-ratio  tests
model_p10q1 <- update(model_p10,
      correlation=corARMA(q=1,p=10,form=~time|nile))
anova(model_p10,model_p10q1)

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

```
> anova(model_p10,model_p10q1)
            Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model_p10       1 19 1755.983 1808.945 -858.9915
model_p10q1     2 20 1756.284 1812.034 -858.1421 1 vs 2 1.698731  0.1925

> anova(model_p10,model_p11)
          Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model_p10     1 19 1755.983 1808.945 -858.9915
model_p11     2 20 1755.303 1811.053 -857.6516 1 vs 2 2.679826  0.1016
```

STEP 9: PLOT THE RESULTS

Time series with control model

For intervention status *j*, for group *k*, at time *t:*

Pre-existing trend in the outcome of interest for the control group

Baseline outcome for the control group

Baseline difference between intervention and control groups

Pre-existing difference in trend between intervention And control group

Level change in the control group

$$outcome_{jkt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot group_k + \beta_3 \cdot group_k \cdot time_t + \beta_4 \cdot level_t + \beta_5 \cdot trend_{jt} + \beta_6 \cdot level_{jt} \cdot group_k + \beta_7 \cdot trend_{jt} \cdot group_k + \varepsilon_{jkt}$$

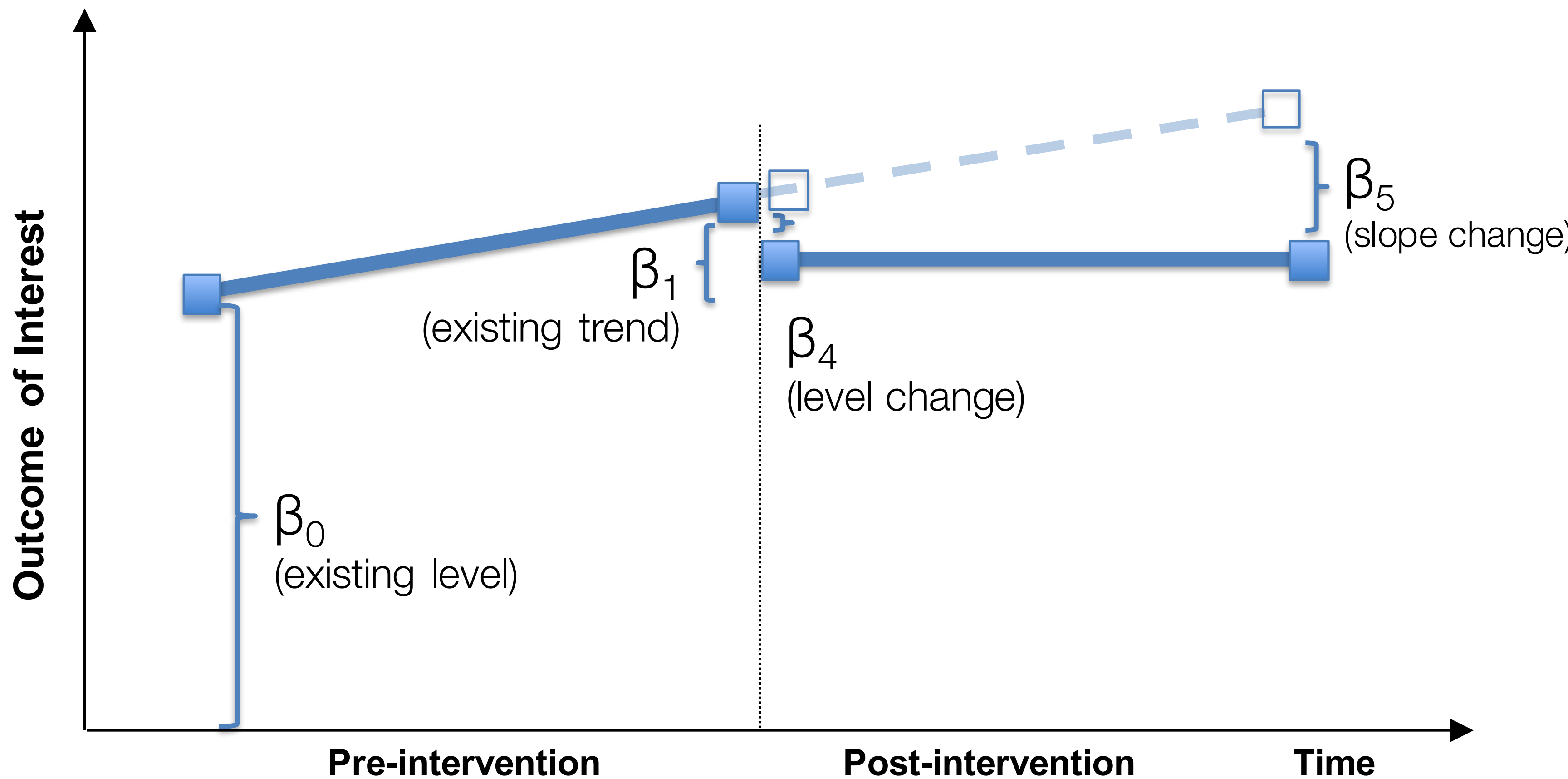
Trend change in the control group

Difference in level change between the intervention and control group
* First variable of interest

Difference in trend change between the intervention and control group
* Second variable of interest

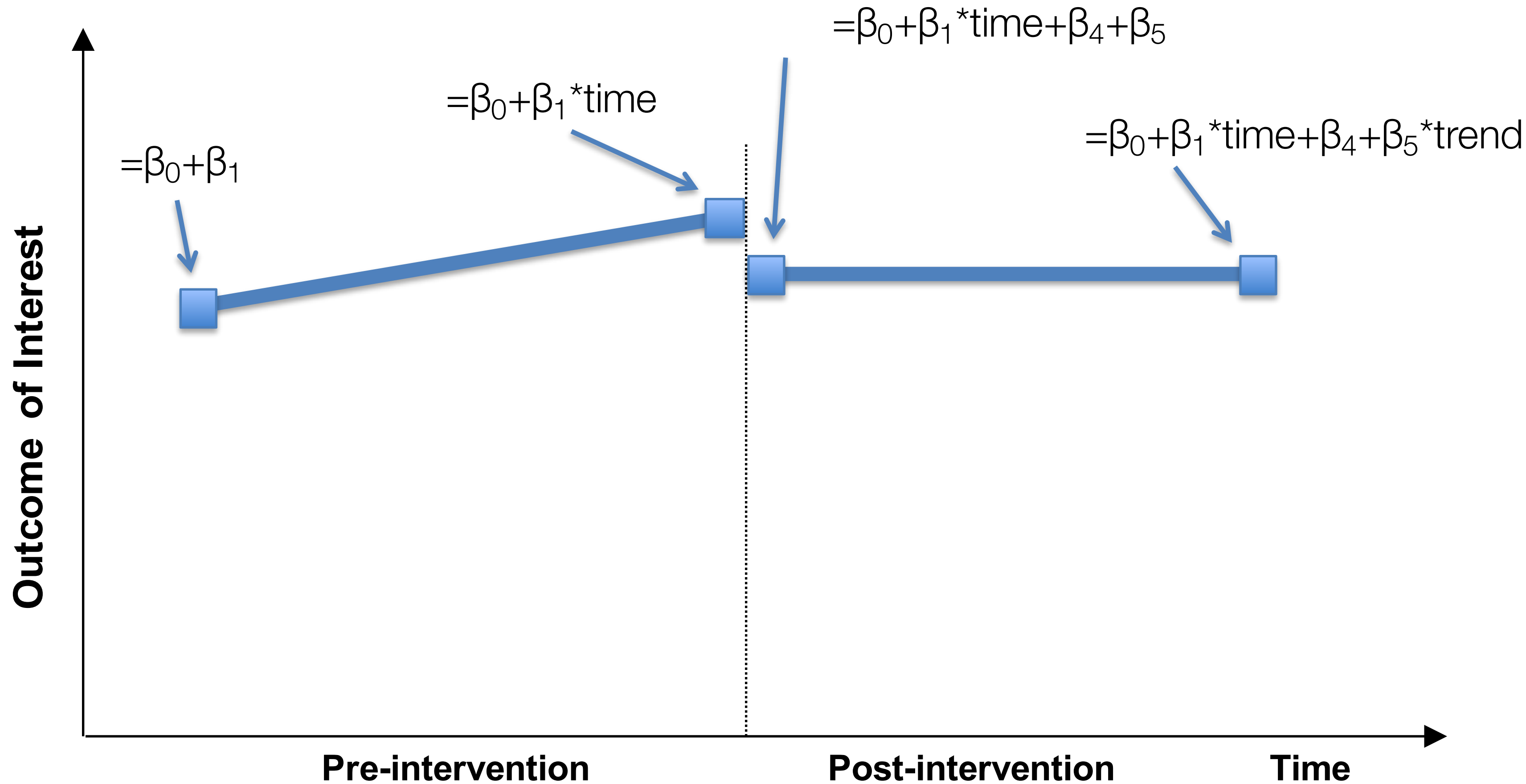
$$outcome_{jkt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot group_k + \beta_3 \cdot group_k \cdot time_t + \beta_4 \cdot level_t + \beta_5 \cdot trend_{jt} + \beta_6 \cdot level_{jt} \cdot group_k + \beta_7 \cdot trend_{jt} \cdot group_k + \varepsilon_{jkt}$$

For the control group:



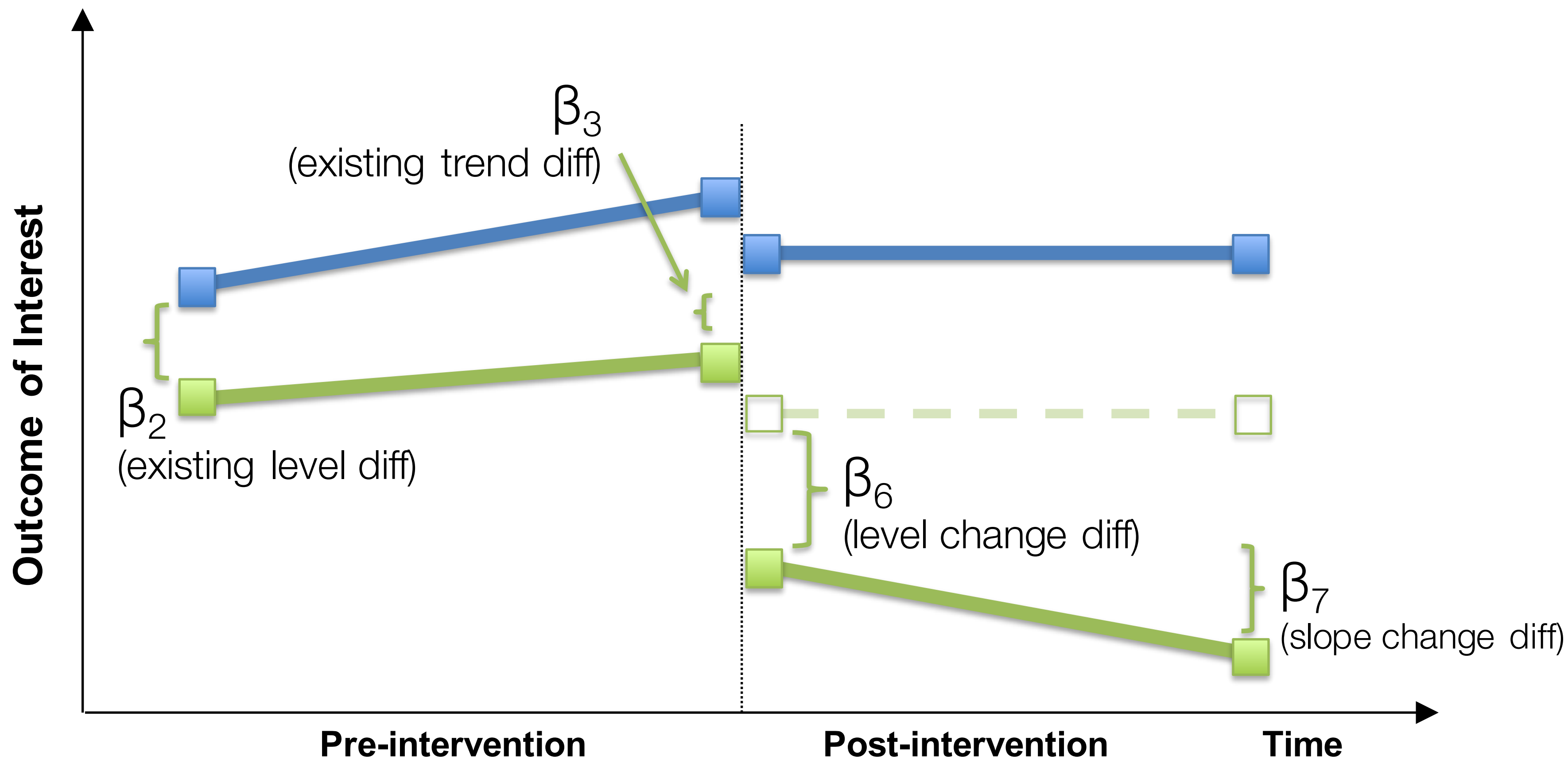
$$outcome_{jkt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot group_k + \beta_3 \cdot group_k \cdot time_t + \beta_4 \cdot level_t + \beta_5 \cdot trend_{jt} + \beta_6 \cdot level_{jt} \cdot group_k + \beta_7 \cdot trend_{jt} \cdot group_k + \varepsilon_{jkt}$$

Predicted values:



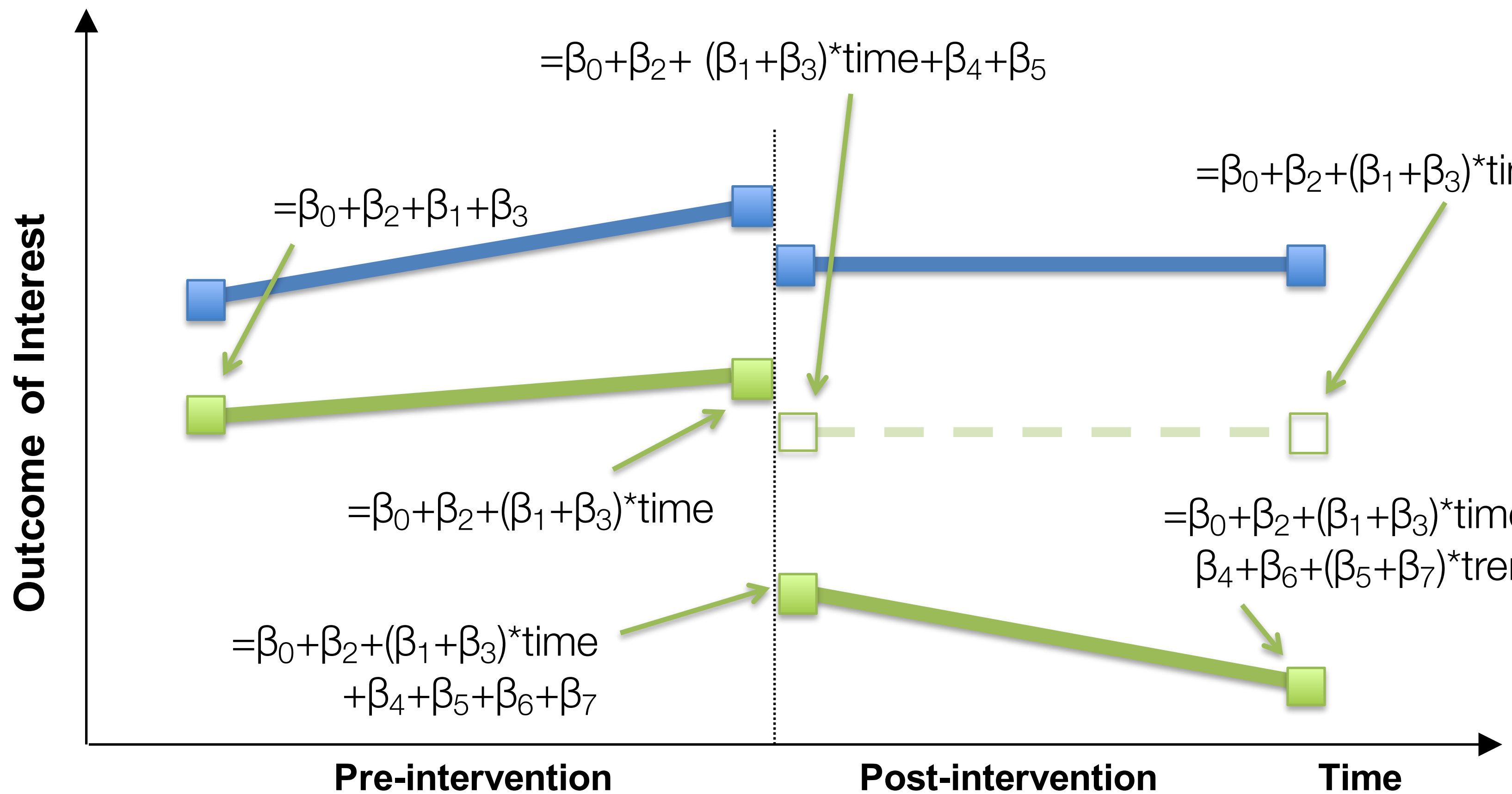
$$outcome_{jkt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot group_k + \beta_3 \cdot group_k \cdot time_t + \beta_4 \cdot level_t + \beta_5 \cdot trend_{jt} + \beta_6 \cdot level_{jt} \cdot group_k + \beta_7 \cdot trend_{jt} \cdot group_k + \varepsilon_{jkt}$$

For the intervention group:

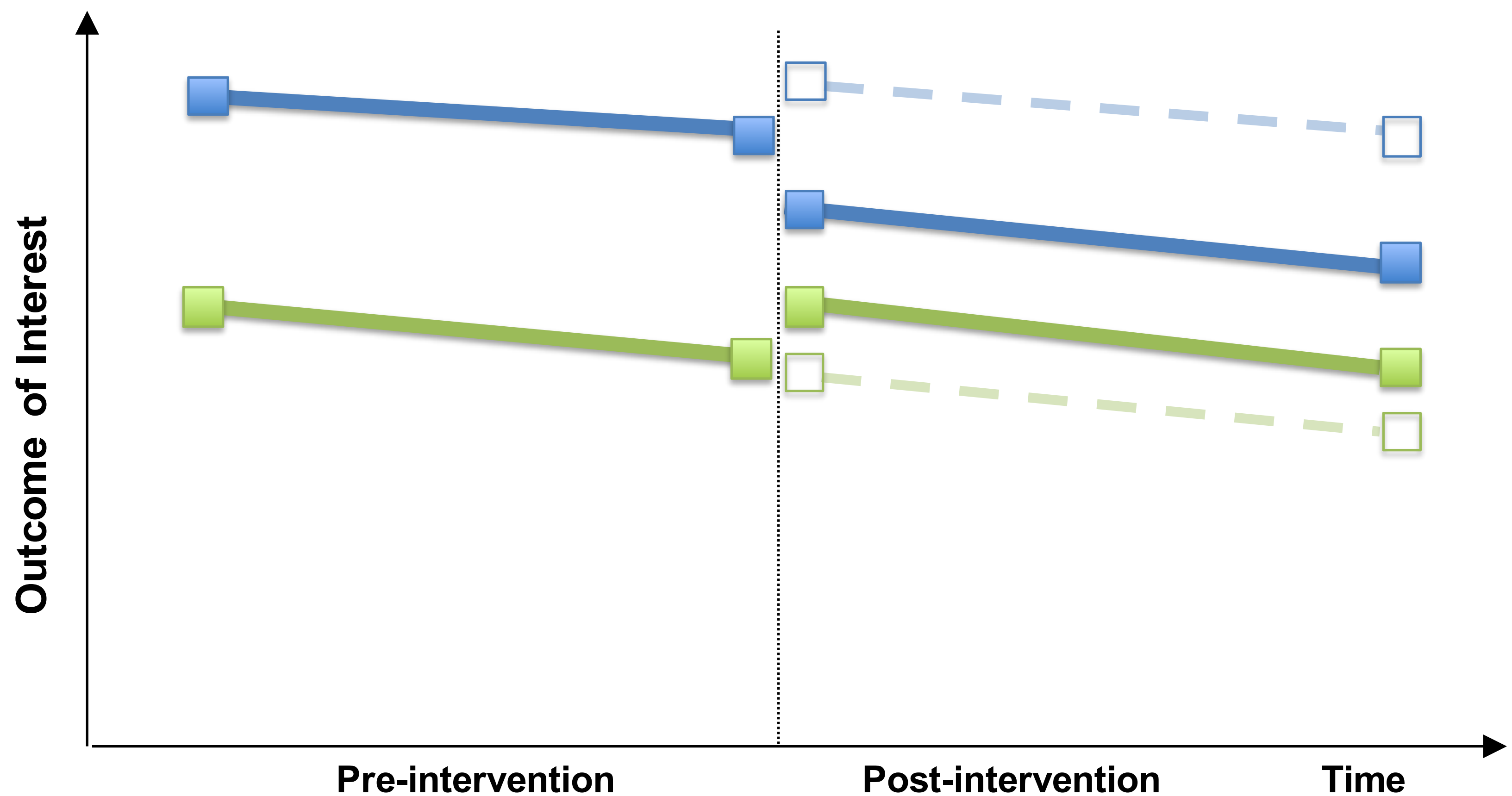


$$outcome_{jkt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot group_k + \beta_3 \cdot group_k \cdot time_t + \beta_4 \cdot level_t + \beta_5 \cdot trend_{jt} + \beta_6 \cdot level_{jt} \cdot group_k + \beta_7 \cdot trend_{jt} \cdot group_k + \varepsilon_{jkt}$$

Predicted values:



```
Coefficients:
(Intercept)  2464.8
time         -20.21
nile          885.4
niletime       26.8
level         368.1
trend           2.1
nilelevel   -1101.2
niletrend     -16.5
```



STEP 9: PLOT THE RESULTS (PART 2)

Plot the results

```
# First plot the raw data points for the Nile
plot(data$time[1:60],data$flow[1:60],
          ylim=c(0,4500),
          ylab="Water Flow",
          xlab="Year",
          pch=20,
          col="lightblue",
          xaxt="n")

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

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

# Add in the points for the control
points(data$time[61:120],data$flow[61:120],
       col="pink",
       pch=20)
```

Plotting the intervention group

- For observed, use *fitted ( )*

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

# Add the second line segment for the intervention group
lines(data$time[28:60], fitted(model_p10)[28:60], col="blue",lwd=2)
```

Plotting the counterfactual

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

```
# Add the counterfactual for the intervention group
segments(28,
   model_p10$coef[1] + model_p10$coef[2]*28 +
   model_p10$coef[3] + model_p10$coef[4]*28 +
   model_p10$coef[5] + model_p10$coef[6],
   60,
   model_p10$coef[1] + model_p10$coef[2]*60 +
   model_p10$coef[3] + model_p10$coef[4]*60 +
   model_p10$coef[5]+model_p10$coef[6]*33,
   lty=2,col='blue',lwd=2)
```

Plotting the control group

- Again, for observed, use *fitted ( )*

```
# Plot the first line segment for the control group
lines(data$time[61:87], fitted(model_p10)[61:87], col="red",lwd=2)

# Add the second line segment for the control
lines(data$time[88:120], fitted(model_p10)[88:120], col="red",lwd=2)
```

Plotting the counterfactual

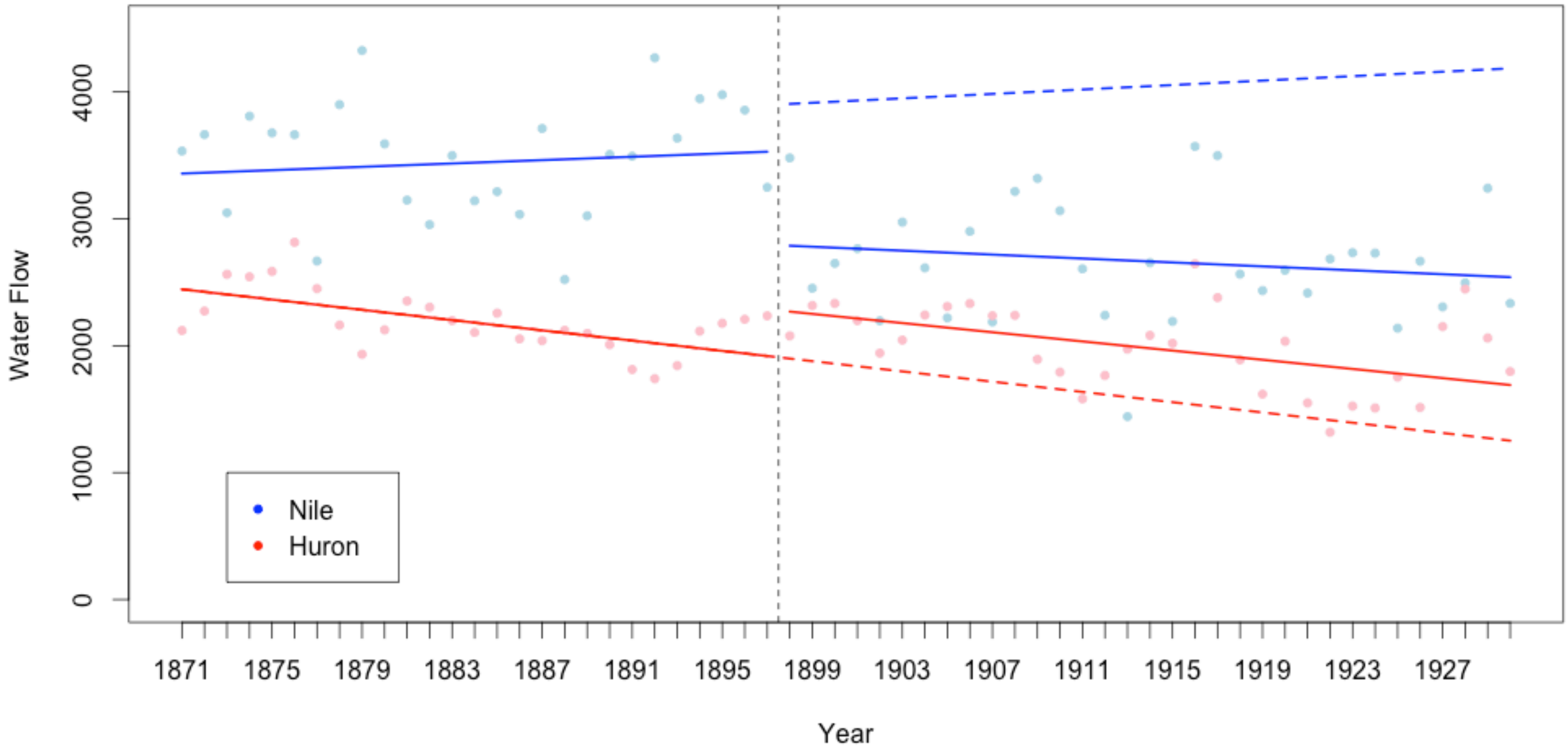
- For counterfactual (if desired), use *segments*(x_1,y_1,x_2,y_2)

```
# Add the counterfactual for the control group
segments(1,
   model_p10$coef[1] + model_p10$coef[2],
   60,
   model_p10$coef[1] + model_p10$coef[2]*60,
   lty=2,col='red',lwd=2)
```

Adding a legend

- Since we have two data series…

```
# Add in a legend
legend(x=3, y=1000,
   legend=c("Nile","Huron"),
   col=c("blue","red"),
   pch=20)
```

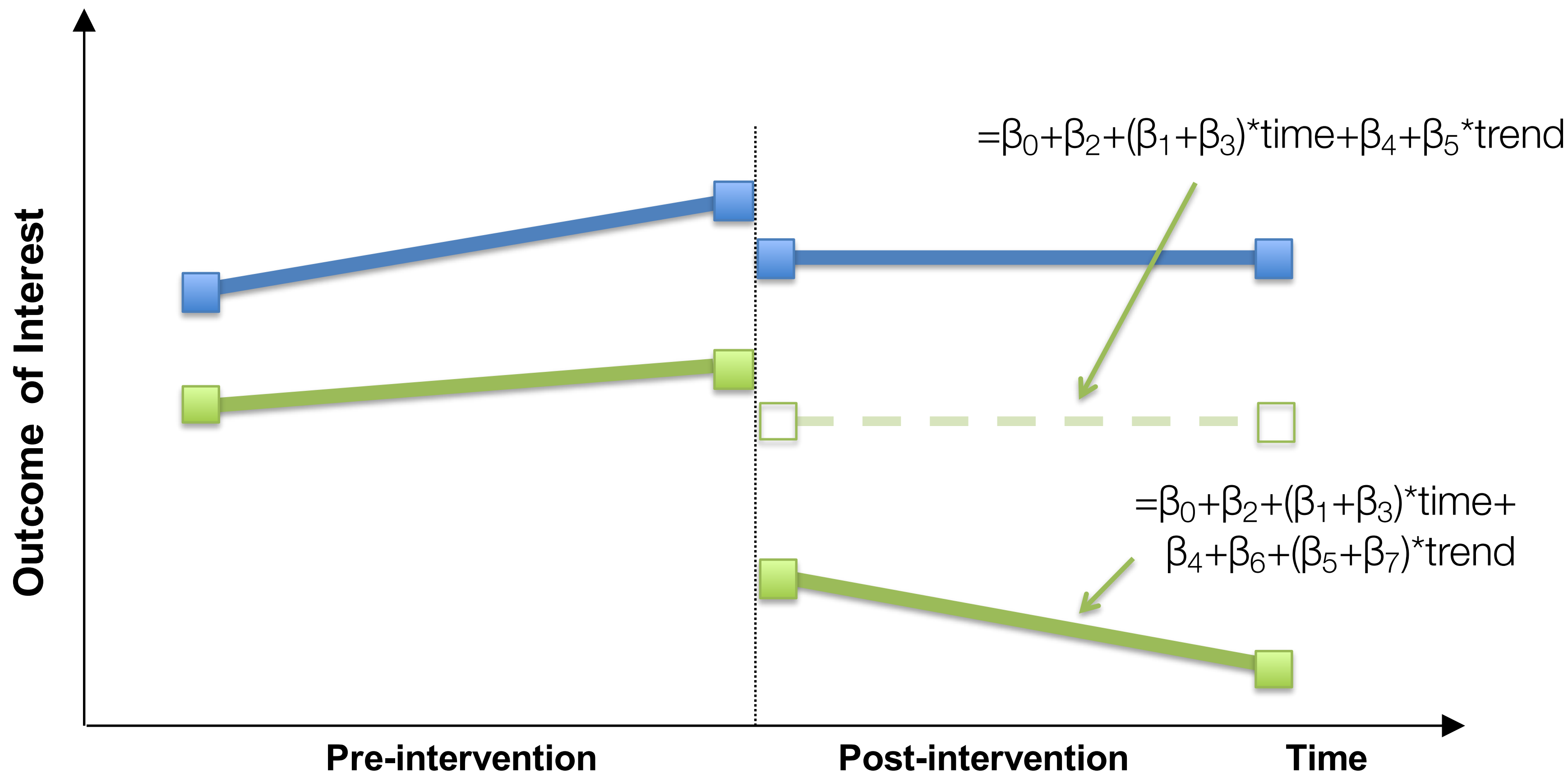


Water Flow
0
1000
2000
3000
4000
Nile
Huron
1871
1875
1879
1883
1887
1891
1895
1899
1903
1907
1911
1915
1919
1923
1927
Year

STEP 10: PREDICTED CHANGES

$$outcome_{jt} = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot group_k + \beta_3 \cdot group_k \cdot time_t + \beta_4 \cdot level_t +$$
$$\beta_5 \cdot trend_t + \beta_6 \cdot level_t \cdot group_k + \beta_7 \cdot trend_t \cdot group_k + \varepsilon_{jkt}$$

Predicted values:



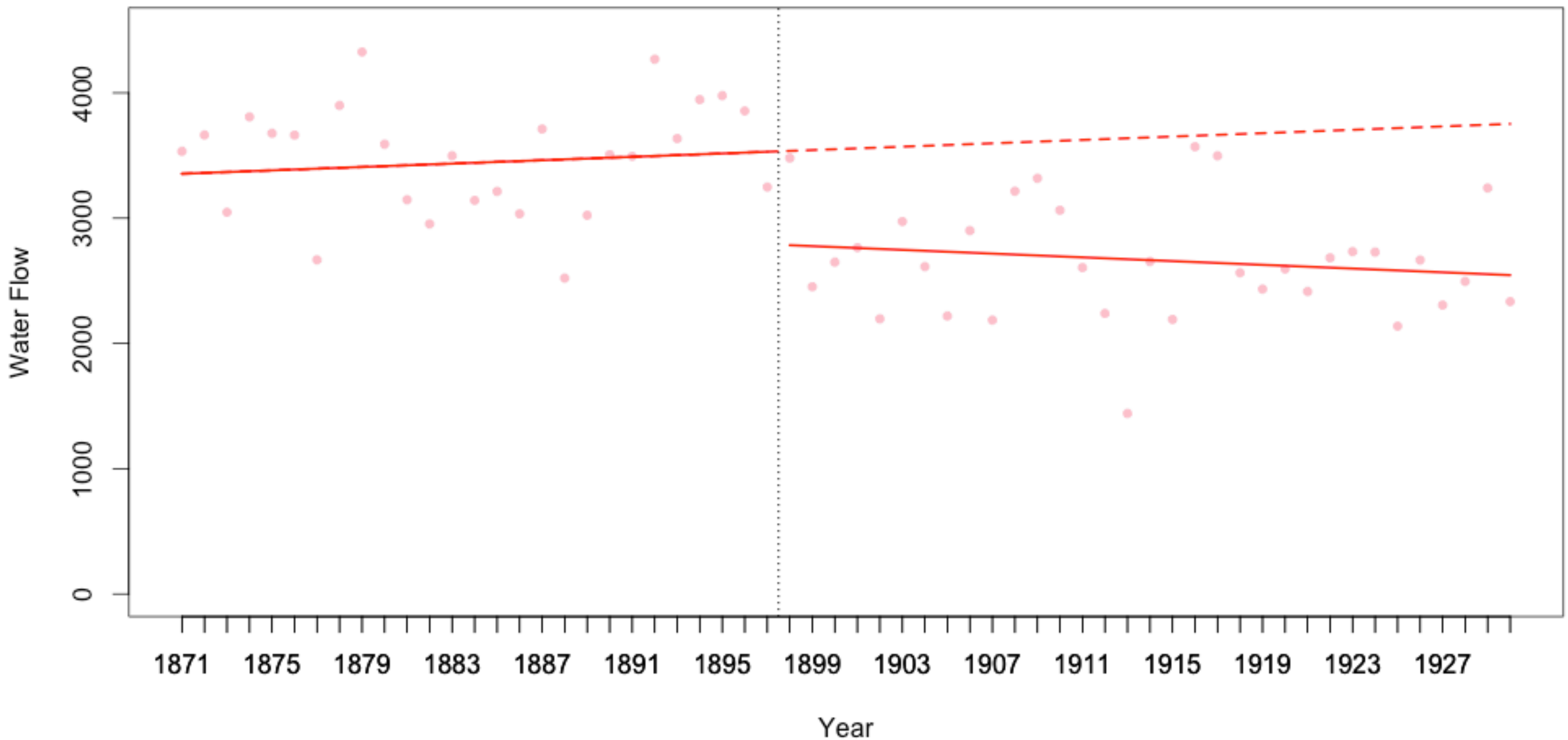
Step 10: Predicted changes

- 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 +
        model_p10$coef[3] + model_p10$coef[4]*52 +
        model_p10$coef[5] + model_p10$coef[6]*25

# Absolute change at 25 years
pred - cfac

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

```
> # Absolute change at 25 years
> pred - cfac
      52
-1513.321

> # Relative change at 25 years
> (pred - cfac) / cfac
       52
-0.3677265
```

Interpretation: In the 25th year after the weather change, the average monthly water flow was 1513 million cubic meters less than would have been expected if the weather had not changed. This represented a 36.8% reduction.



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

	Absolute	Relative
Single Series	-1093	-29.6%
With Control	-1513	-36.8%

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