Australian National University

# Interrupted time series

Alice Richardson

National Centre for Epidemiology & Population Health

16 August 2019

# Overview

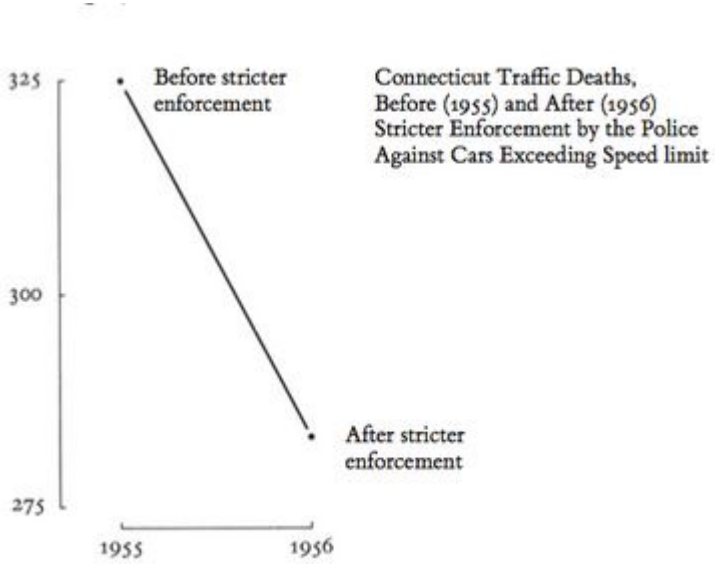
- Introduction
- Examples
- Stata
- R
- Extensions

# Why use an interrupted time series?

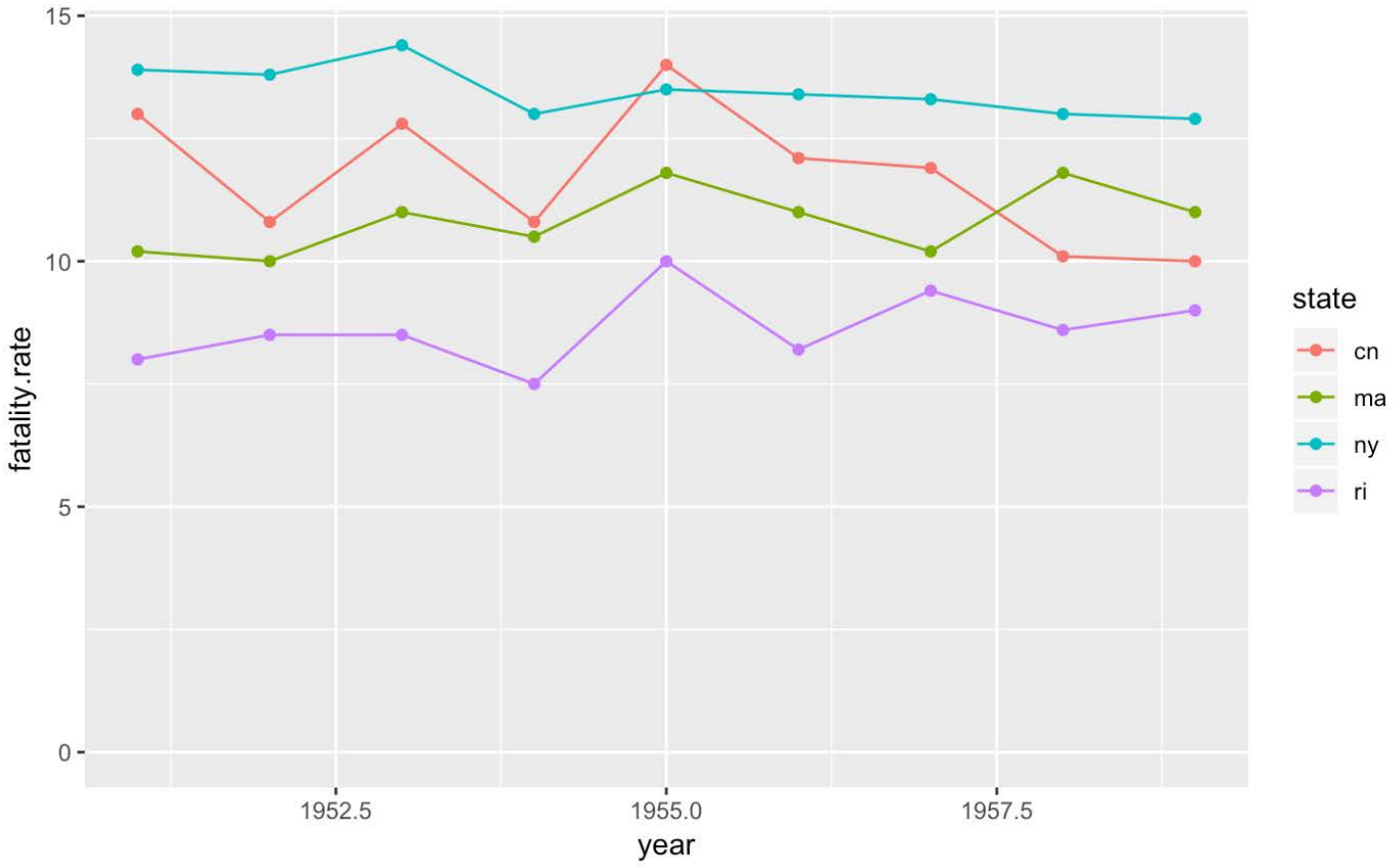
- A randomised experiment is not practical but a quasi-experiment is
- Data collected at equally-spaced time intervals
- Series “interrupted” by an intervention
- Measure effect of the intervention
- Control for time trends

# Examples

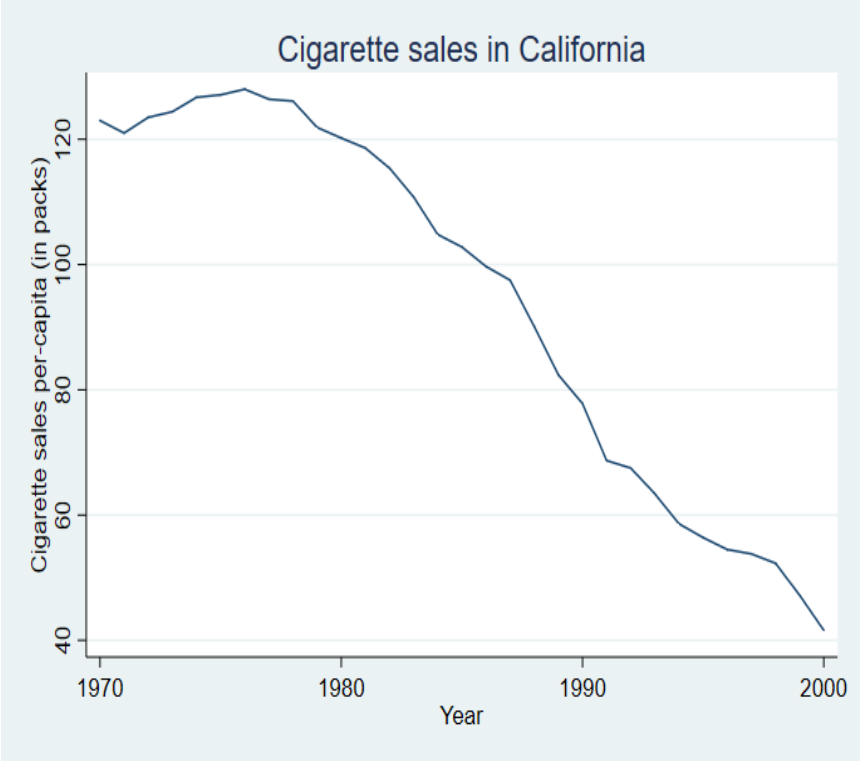
- Speeding in Connecticut: did the introduction of stricter enforcement reduce fatalities?



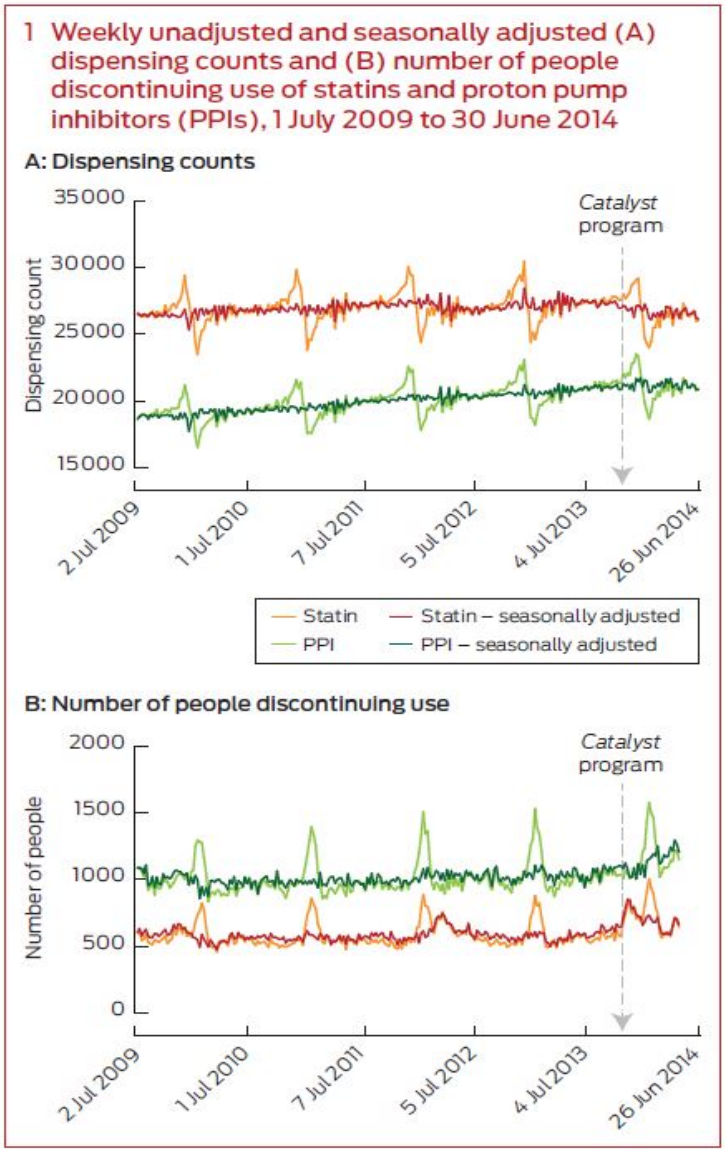

325
300
275
250
cn.deaths
1952.5
1955.0
1957.5
year


15
10
5
0
fatality.rate
1952.5
1955.0
1957.5
year
state
cn
ma
ny
ri

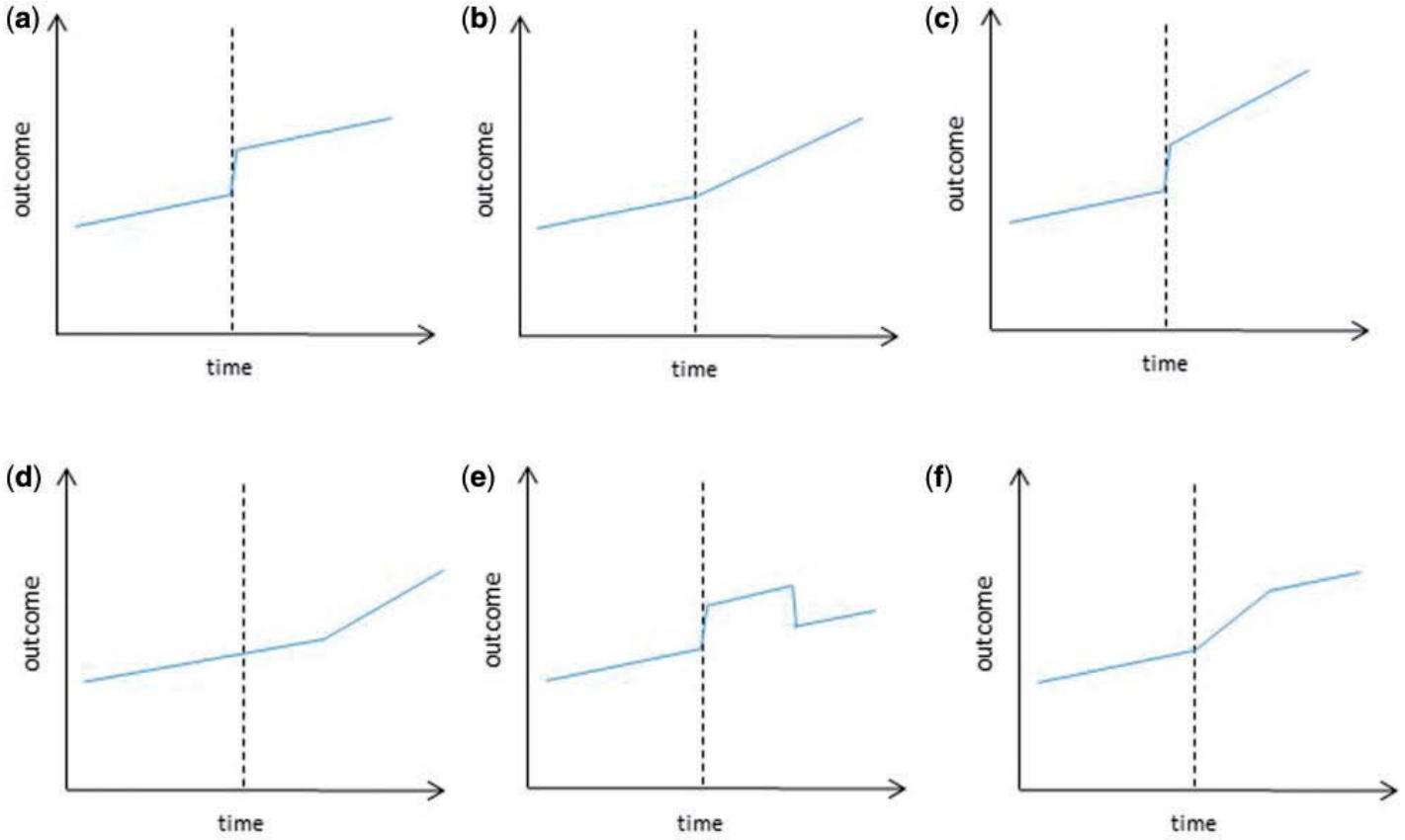
- Cigarette sales: did a tax increase beginning in 1989 reduce sales?



- Catalyst and statin sales: did the TV show influence people to stop taking their statins?



# Types of impact




- Level change; Slope change; Level and slope change; Slope change with a lag; Temporary level change; Temporary slope change leading to a level change

# Stata: set up and fit model

```
.  use http://fmwww.bc.edu/repec/bocode/c/cigsales.dta

.
.  tsset state year
        panel variable:  state (strongly balanced)
         time variable:  year, 1970 to 2000
                 delta:  1 unit

.
.  itsa cigsale, single treat(3) trperiod(1989) lag(1) posttrend figure


        panel variable:  state (strongly balanced)
         time variable:  year, 1970 to 2000
                 delta:  1 unit
```

Regression with Newey-West standard errors
maximum lag: 1

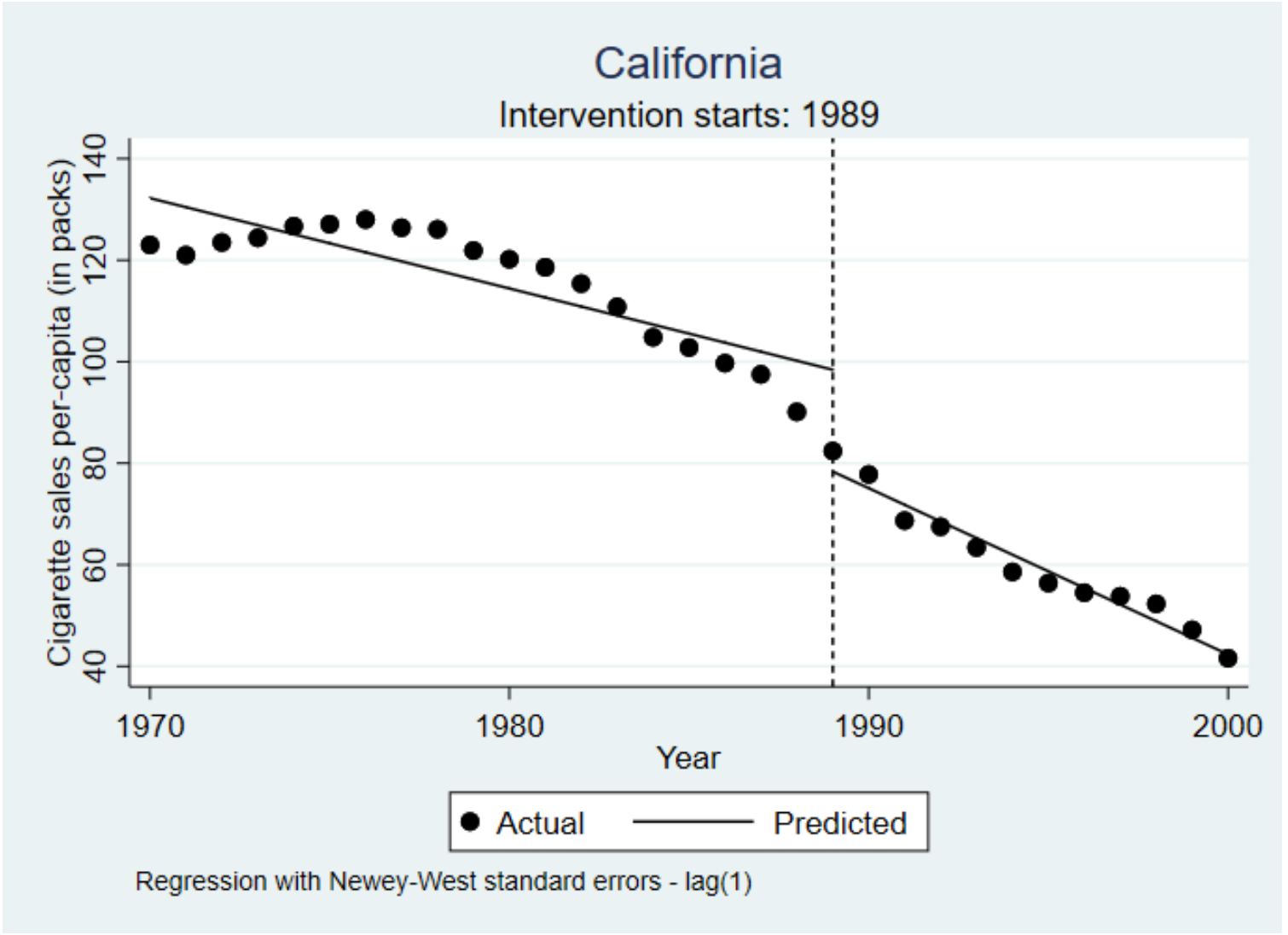
Number of obs = 31
F( 3, 27) = 331.45
Prob > F = 0.0000

| cigsale | Coef. | Newey-West Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| _t | -1.779474 | .3834188 | -4.64 | 0.000 | -2.566184 | -.9927632 |
| _x1989 | -20.0581 | 4.724395 | -4.25 | 0.000 | -29.75175 | -10.36444 |
| _x_t1989 | -1.494652 | .4368201 | -3.42 | 0.002 | -2.390933 | -.5983715 |
| _cons | 132.2258 | 4.253054 | 31.09 | 0.000 | 123.4992 | 140.9523 |

Postintervention Linear Trend: 1989

Treated: _b[_t]+_b[_x_t1989]

| Linear Trend | Coeff | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Treated | -3.2741 | 0.2688 | -12.1803 | 0.0000 | -3.8257 | -2.7226 |


California
Intervention starts: 1989
Cigarette sales per-capita (in packs)
40
60
80
100
120
140
1970
1980
1990
2000
Year
Actual
Predicted
Regression with Newey-West standard errors - lag(1)

# Test for autocorrelation

```
. actest, lag(6)

Cumby-Huizinga test for autocorrelation
  H0: variable is MA process up to order q
  HA: serial correlation present at specified lags >q
```

| H0: q=0 (serially uncorrelated) HA: s.c. present at range specified | | | | H0: q=specified lag-1 HA: s.c. present at lag specified | | | |
|---|---|---|---|---|---|---|---|
| lags | chi2 | df | p-val | lag | chi2 | df | p-val |
| 1 - 1 | 15.242 | 1 | 0.0001 | 1 | 15.242 | 1 | 0.0001 |
| 1 - 2 | 15.255 | 2 | 0.0005 | 2 | 3.300 | 1 | 0.0693 |
| 1 - 3 | 15.325 | 3 | 0.0016 | 3 | 1.192 | 1 | 0.2749 |
| 1 - 4 | 15.896 | 4 | 0.0032 | 4 | 0.000 | 1 | 0.9880 |
| 1 - 5 | 16.057 | 5 | 0.0067 | 5 | 1.113 | 1 | 0.2914 |
| 1 - 6 | 16.078 | 6 | 0.0133 | 6 | 2.051 | 1 | 0.1521 |

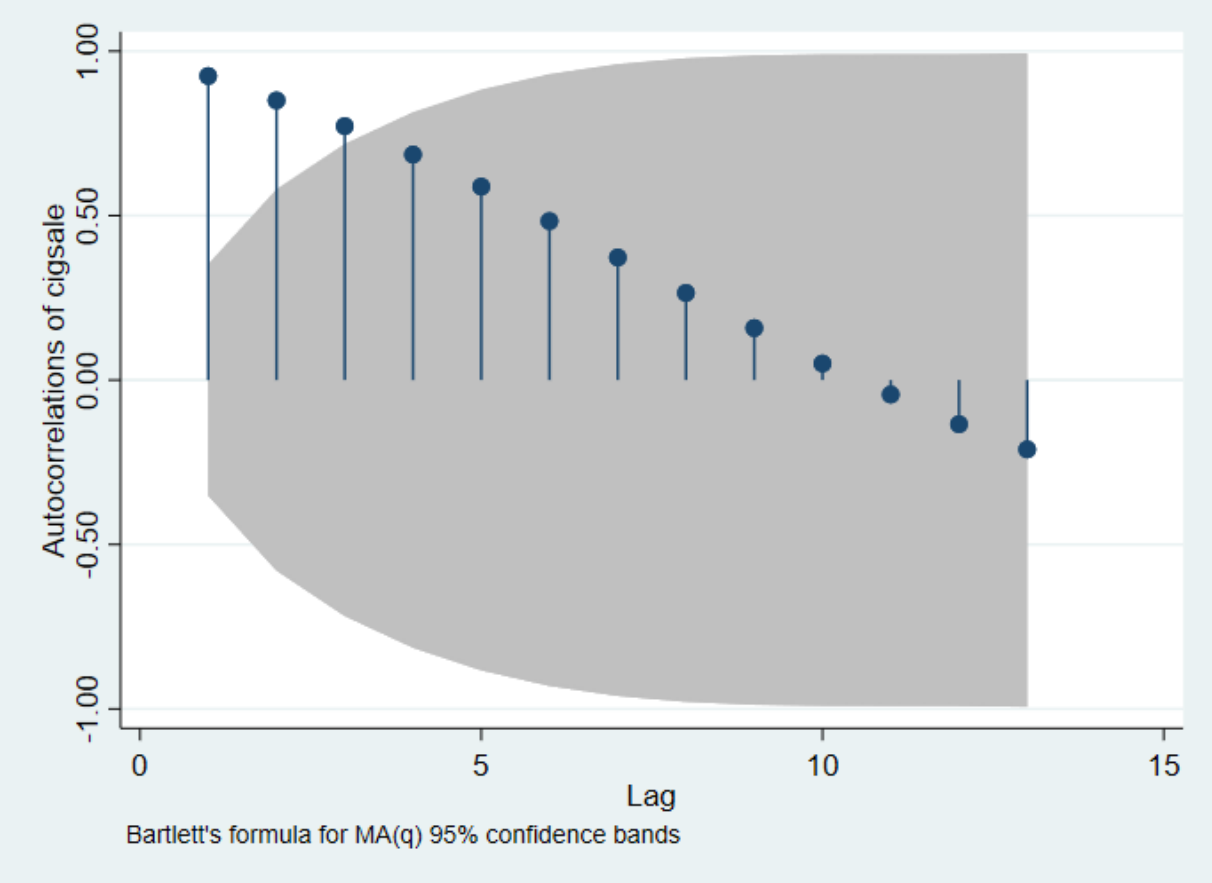
```
  Test allows predetermined regressors/instruments
  Test requires conditional homoskedasticity
```

# Autocorrelation plots

```
. ac cigsale

. corrgram cigsale

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.9246   1.0380   29.154  0.0000          |-------           |--------
2        0.8504  -0.4160   54.665  0.0000          |------         ---|
3        0.7721  -0.2780   76.444  0.0000          |------          --|
4        0.6855   0.0421   94.248  0.0000          |-----             |
5        0.5881   0.1094   107.86  0.0000          |----              |
6        0.4834  -0.2862   117.42  0.0000          |---            --|
7        0.3727  -0.0382   123.34  0.0000          |--                |
8        0.2649   0.2119   126.46  0.0000          |--                |-
9        0.1582   0.0841   127.62  0.0000          |-                 |
10       0.0504  -0.3709   127.75  0.0000          |                --|
11      -0.0441  -0.1241   127.85  0.0000          |                  |
12      -0.1343  -0.4970   128.82  0.0000         -|               ---|
13      -0.2109  -0.7565   131.35  0.0000         -|            ------|
```


Autocorrelations of cigsale
1.00
0.50
0.00
-0.50
-1.00
0
5
10
15
Lag
Bartlett's formula for MA(q) 95% confidence bands

# Alternatively, model AR(1) directly

- `itsa cigsale, single treat(3) trperiod(1989) replace prais rhotype(tscorr) vce(robust)`

Prais-Winsten AR(1) regression -- iterated estimates

Linear regression

| | |
|---|---|
| Number of obs | = 31 |
| F(3, 27) | = 609.24 |
| Prob > F | = 0.0000 |
| R-squared | = 0.9011 |
| Root MSE | = 2.5964 |

| cigsale | Coef. | Semirobust Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| _t | -1.843139 | .4538631 | -4.06 | 0.000 | -2.77439 | -.9118892 |
| _x1989 | -6.094491 | .8840197 | -6.89 | 0.000 | -7.90835 | -4.280633 |
| _x_t1989 | -1.998494 | .9191 | -2.17 | 0.039 | -3.884332 | -.1126568 |
| _cons | 126.35 | 3.789489 | 33.34 | 0.000 | 118.5746 | 134.1254 |
| rho | .9424635 | | | | | |

Durbin-Watson statistic (original) 0.535242
Durbin-Watson statistic (transformed) 1.342728

# R set up and fit model

- `data$smokban <- c(rep(0, 19), rep(1, nrow(data) - 19))`
- `data$year1970 <- data$year - 1970`
- `cigsale.model01 <- lm(cigsale ~ smokban*year1970, data = data)`
- `summary(cigsale.model01)`

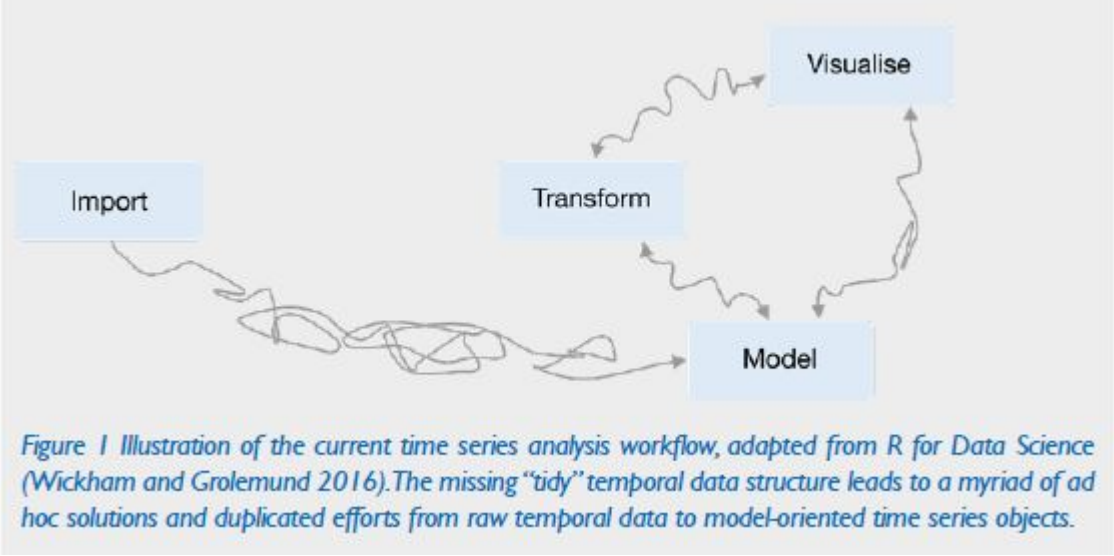
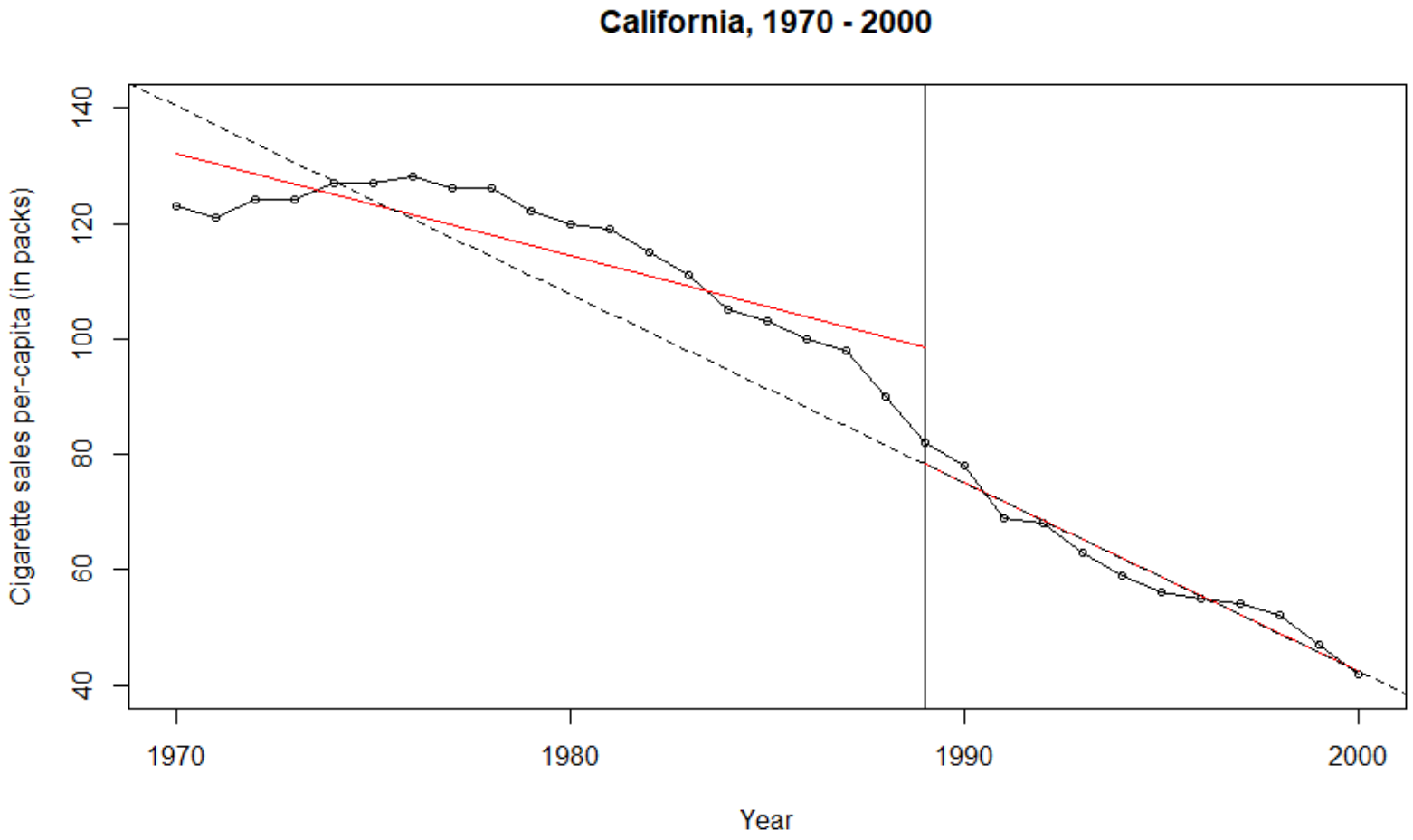



*Figure 1 Illustration of the current time series analysis workflow, adapted from R for Data Science (Wickham and Grolemund 2016). The missing "tidy" temporal data structure leads to a myriad of ad hoc solutions and duplicated efforts from raw temporal data to model-oriented time series objects.*

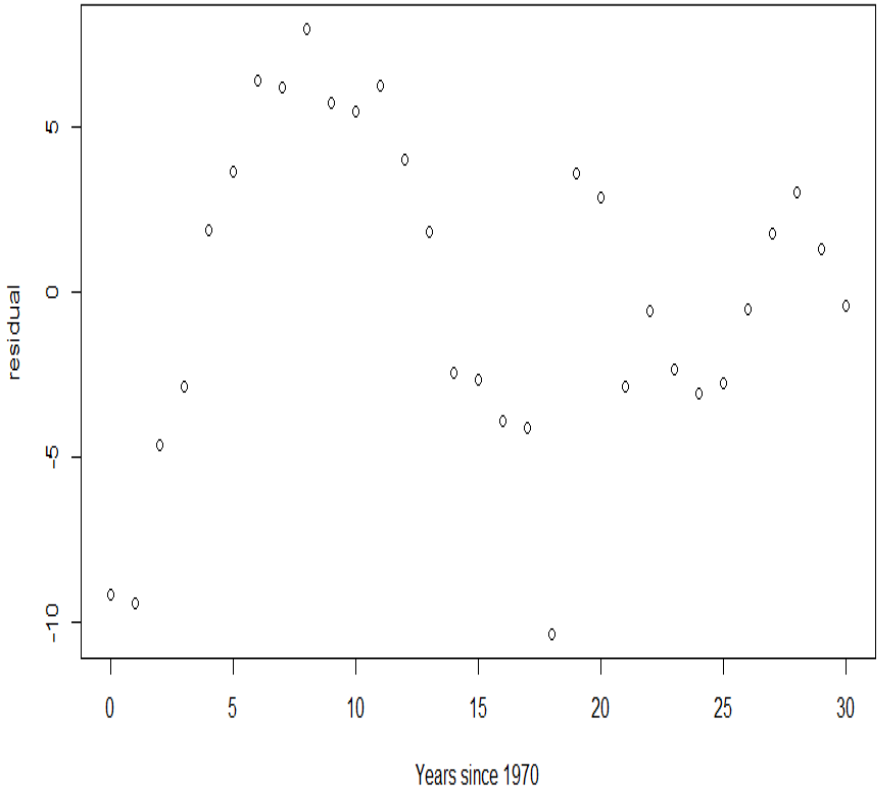
```
Call:
lm(formula = cigsale ~ smokban * year1970, data = data)

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      132.1789     2.2528  58.673  < 2e-16 ***
smokban            8.3339    10.8001   0.772  0.44703
year1970          -1.7684     0.2138  -8.270 7.05e-09 ***
smokban:year1970  -1.5008     0.4775  -3.143  0.00403 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.105 on 27 degrees of freedom
Multiple R-squared:  0.9739,       Adjusted R-squared:  0.971
F-statistic: 335.7 on 3 and 27 DF,  p-value: < 2.2e-16
```


California, 1970 - 2000
Cigarette sales per-capita (in packs)
140
120
100
80
60
40
1970
1980
1990
2000
Year

- dwtest(cigsale.model01)
- Durbin-Watson test
- data: cigsale.model01
- DW = 0.55595, p-value = 8.806e-09
- alternative hypothesis: true autocorrelation is greater than 0



# Model AR(1) directly

- Generalized least squares fit by REML
- Model: cigsale ~ smokban * year1970

- Correlation Structure: AR(1). Formula: ~1
- Parameter estimate(s): Phi = 0.999993

- Coefficients:

| | Value | Std.Error | t-value | p-value |
|---|---|---|---|---|
| (Intercept) | 123.00058 | 719.2730 | 0.1710068 | 0.8655 |
| smokban | 28.09091 | 19.3476 | 1.4519066 | 0.1580 |
| year1970 | -1.83333 | 0.6335 | -2.8940451 | 0.0074 |
| smokban:year1970 | -1.80303 | 1.0286 | -1.7528736 | 0.0910 |

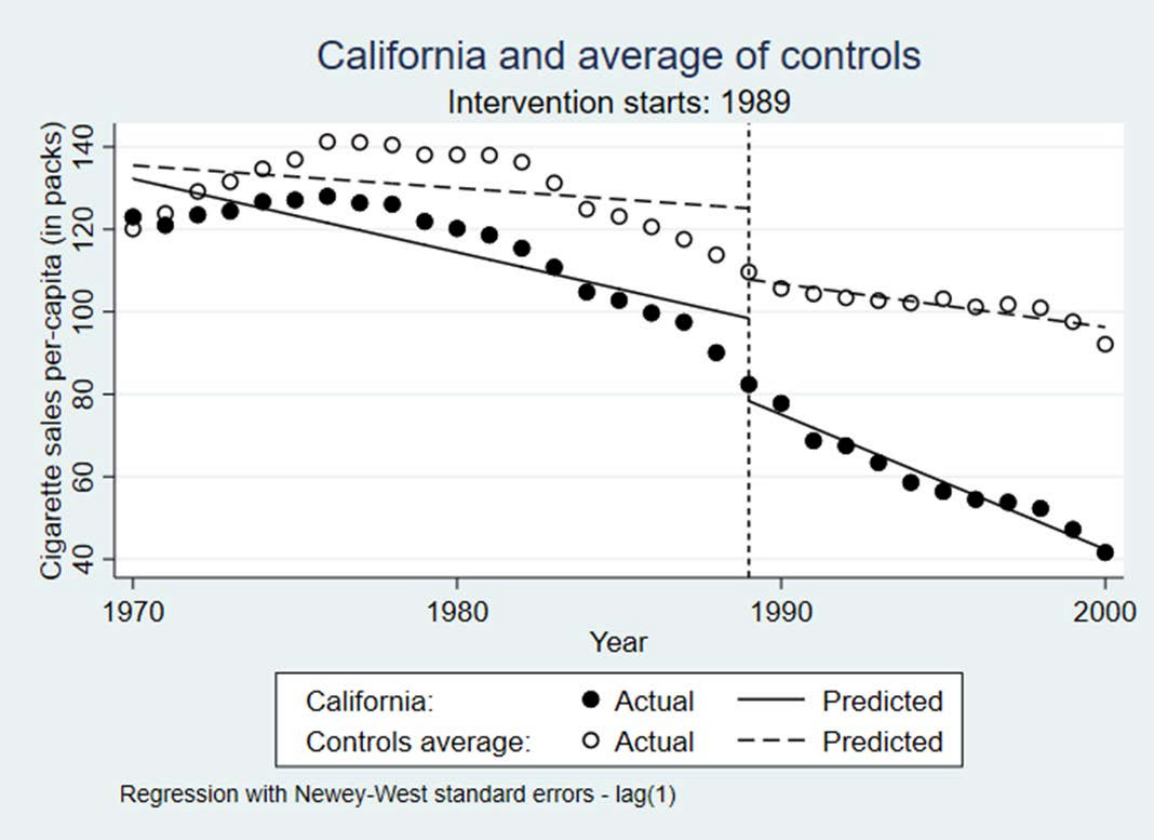
- Residual standard error: 719.273
- Degrees of freedom: 31 total; 27 residual

# Issues

- How many observations?
  - Three?
  - Eight?

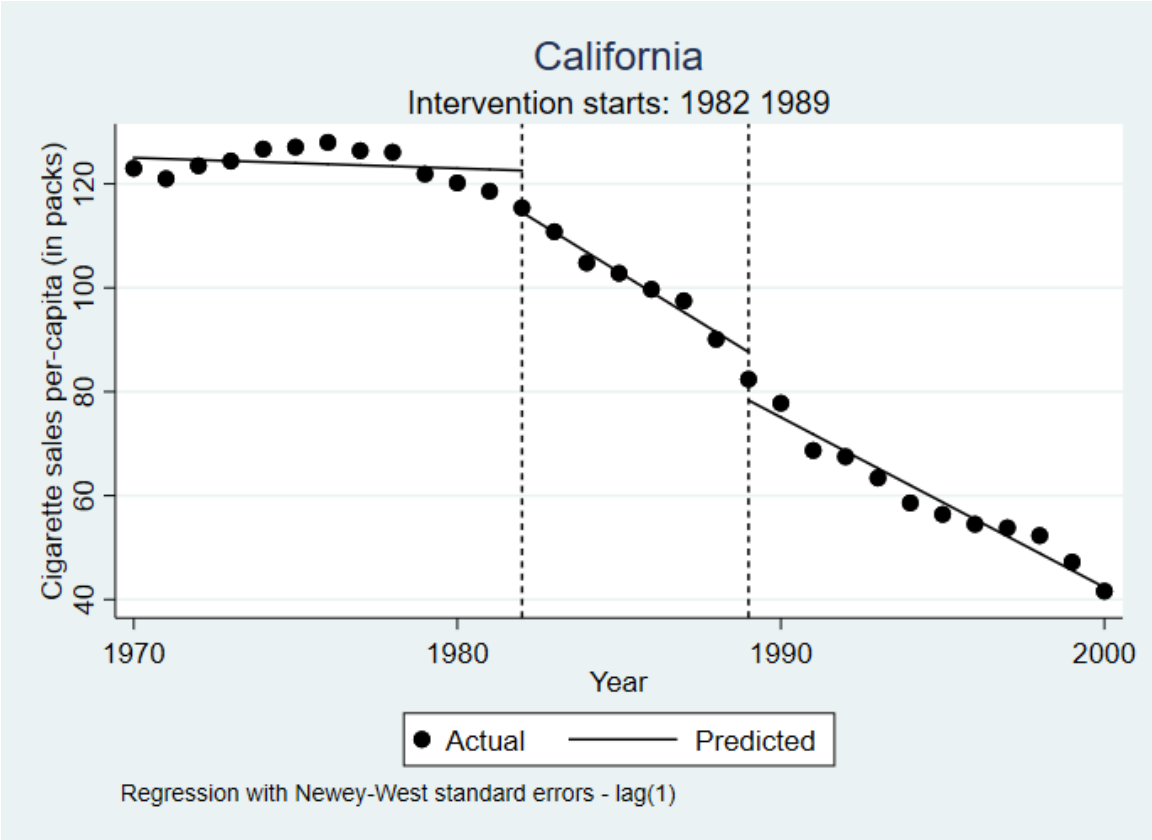
# Extensions

- Multiple-group design
- Compare intervention group with control group(s)

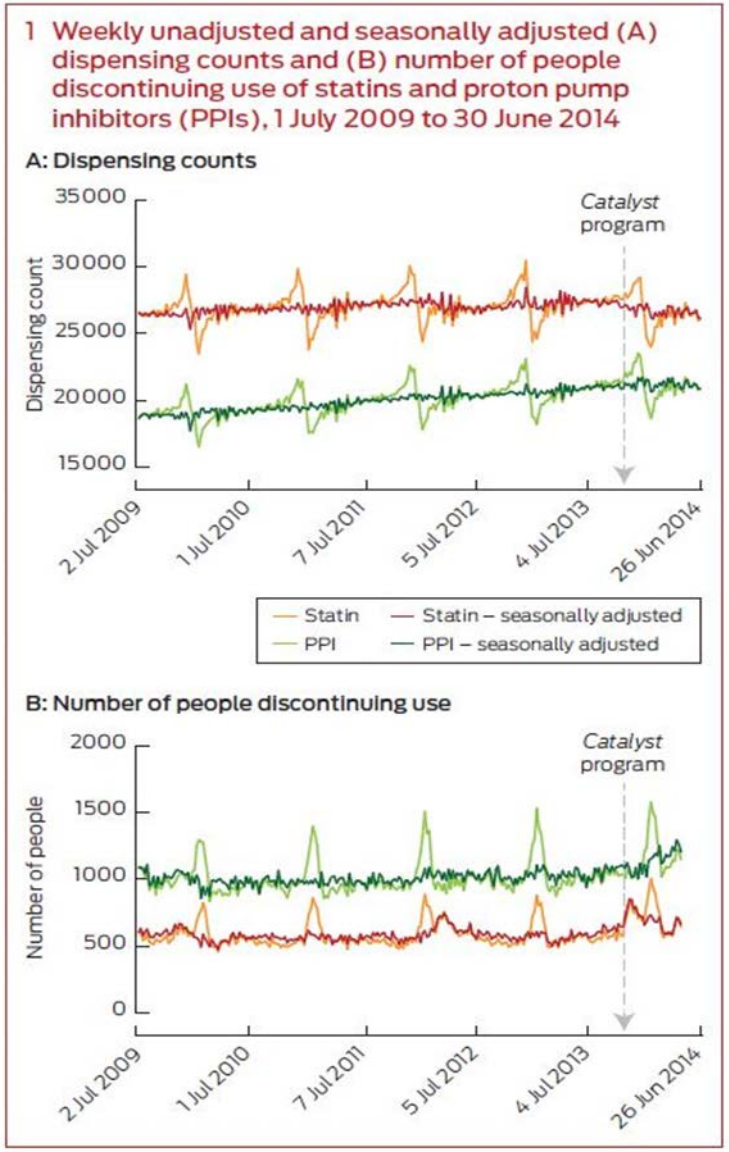


- Types of control group
  - Location e.g. state
  - Characteristic e.g. gender
  - Behaviour e.g. age group
  - Historical e.g. time periods
  - Outcome e.g. injury vs mortality
  - Time-period e.g. time of day

- Intervention starts and stops; or
- Multiple interventions



- Seasonality

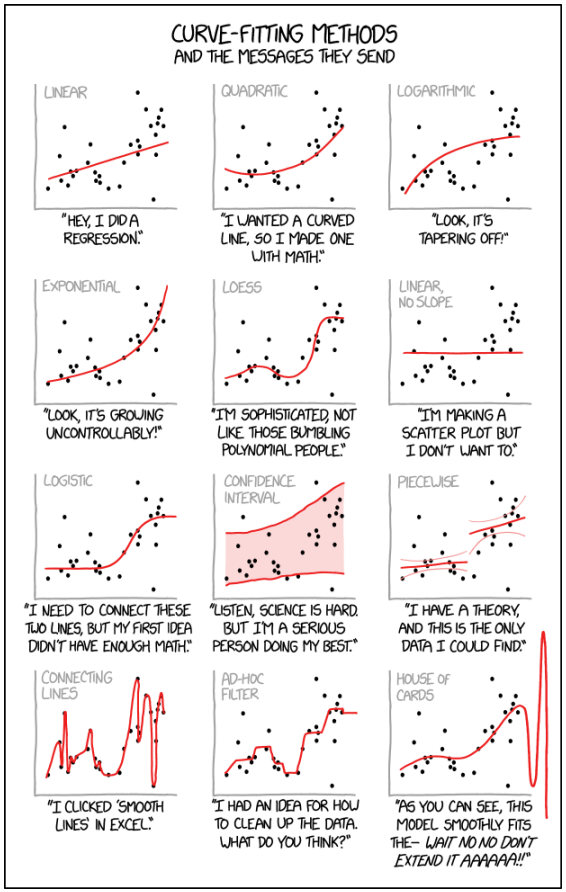


- Extra variability
  - Increasing variance?
  - Count outcomes with Poisson distribution?
  - Model directly, not using itsa
- Time-varying confounders?
  - Include as covariates in regression model

# Resources

- IDRE: nothing I could find
- Alternative names include changepoint analysis and piecewise or segmented regression

# Parting shot ...



# References


- Bernal et al (2017). Interrupted time series regression for the evaluation of public health interventions: a tutorial. IJE 46, 348 – 355.
- Bernal et al (2018). The use of controls in interrupted time series studies of public health interventions. IJE 47, 2082 – 2093.
- Bernal et al (2018). A methodological framework for model selection in interrupted time series studies. J Clin Epi 103, 82 – 91.
- Biglan et al. (2000). The value of interrupted time-series experiments for community intervention research. Prevention Science 1, 31-49.
- Schaffer et al. (2015). The crux of the matter: did the ABC's Catalyst program change statin use in Australia? MJA 202, 591 – 595.
- Linden (2015). Conducting interrupted time-series analysis for single- and multiple-group comparisons. Stat Journal 15, 480 – 500.

# Upcoming R workshops
# 10am-12, Tony McMichael seminar room

| Date | Topic |
|---|---|
| 21 August | Drop-in R setup |
| 22 August | Packages, importing data, scripts |
| 5 September | Visualisation (base R) |
| 19 September | Visualisation (ggplot) |
| 26 September | Meta analysis (for EPP) |
| 3 October | R Markdown |
| 17 October | Logistic regression modelling |
| 31 October | Power and sample size |