ALR 2014 Caual Modelling Workshop

Example 2: Difference in Differences

Do Medical Marijuana Laws Increase Marijuana Use? Replication Study and Extension

[Sam Harper, Erin C Strumpf, and Jay S Kaufman. Do Medical Marijuana Laws Increase Marijuana Use? Replication Study and Extension. Annals of Epidemiology. 2012;22(4):207â€“212.]

Analysis set up

Set working directory and importing the data into R

```
options(scipen = 1, digits = 2)
opts_chunk$set(warning = FALSE, message = FALSE, error = FALSE)

setwd("/Users/")
mml_data <- read.csv("http://www.walkabilly.net/Presentations/Workshops/mml_data.csv")
```

Installing necessary packages

The downloaded binary packages are in /var/folders/gw/_z_ntf0x0wn_jhmq2rb23jy00000gn/T//RtmpaSXP8k/downloaded_packages The downloaded binary packages are in /var/folders/gw/_z_ntf0x0wn_jhmq2rb23jy00000gn/T//RtmpaSXP8k/downloaded_packages The downloaded binary packages are in /var/folders/gw/_z_ntf0x0wn_jhmq2rb23jy00000gn/T//RtmpaSXP8k/downloaded_packages The downloaded binary packages are in /var/folders/gw/_z_ntf0x0wn_jhmq2rb23jy00000gn/T//RtmpaSXP8k/downloaded_packages The downloaded binary packages are in /var/folders/gw/_z_ntf0x0wn_jhmq2rb23jy00000gn/T//RtmpaSXP8k/downloaded_packages

What is the impact of legalizing medical marijuana on recreational marijuana use?

Some states do and some states do not have medical marijuana laws (MML). As time passes more states have medical marijuana laws. We want to estimate the impact of implementing these laws on recreational marijuana use.

A key confounder we are trying to control is the fact that states that pass MMLs may differ from those that do not in ways that may also be correlated with marijuana use. For example, if states passing laws tend to have more liberal social norms about drug use, this could be mistaken as the effect of the policy.

This is a very similar problem to the residential self-selection problem in active living research. Refrased, we could say that neighbourhoods that implement different active living friend policies (i.e., parks, trafffic calming, cycling infrastructure) may differ from those that do not in ways that may also be correlated with active living.

Replication of past research

[Wall, M., Poh, E., CerdÃ¡, M., Keyes, K. M., Galea, S., Hasin, D. S. Adolescent marijuana use from 2002 to 2008: higher in states with medical marijuana laws, cause still unclear. Annals of Epidemiology. 2011;21(9):714â€“716.]

Subset the data to only include 2002 to 2007 data

```r
mml_data <- subset(mml_data, mml_data$year >= 2002 & mml_data$year <= 2007)
```

Recode the main dummy variable in the regression.

```r
library(car)
mml_data$r_wall <- recode(mml_data$wall, "'Never MML'=0; 'MML-pre'=1; 'MML-post'=2;",
    as.factor.result = TRUE)
```

Subset the data to 12 to 17 years old youth only

```r
mml_data_12to17 <- subset(mml_data, mml_data$age == "12-17y")
```

Mean and standard deviation of past month marijuana use rate (%)

```r
library(doBy)
library(xtable)
treatmean <- summaryBy(pmu ~ treatex, data = mml_data_12to17, FUN = function(x,
    na.rm = FALSE) {
    c(m = mean(x), s = sd(x))
})
print(xtable(treatmean), type = "html")
```

	treatex	pmu.m	pmu.s
1	no	6.94	1.21
2	yes	8.68	1.67

Regression results from Wall et al., 2011

Regression equation 1

$$Y_{st} = \beta_0 + \beta_1 MML_{st} + \gamma_t + \varepsilon_{st}$$

- Y(st) = Marijuana use in state s in year t
- MML(st) = A dummy variable indicating whether or not a state had a MML in place in year t
- Gamma(t) = A fixed effect for each year
- Epsilon(st) = A state-year-specific error term.

Under equation 1, if MMLs were randomly assigned to some states in any given year, b1 would estimate the causal effect of the law on marijuana use.

```r
replication1 <- lm(pmu ~ factor(mmlf) + year0, data = mml_data_12to17, na.action = na.exclude)
summary(replication1)
```

```
##
## Call:
## lm(formula = pmu ~ factor(mmlf) + year0, data = mml_data_12to17,
##     na.action = na.exclude)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.949 -0.907 -0.156  0.705  5.088
##
```

```
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)          8.232      0.179   46.02  < 2e-16 ***
## factor(mmlf)yes      1.871      0.178   10.52  < 2e-16 ***
## year02003-4         -0.543      0.250   -2.17    0.031 *
## year02004-5         -1.124      0.250   -4.50  9.9e-06 ***
## year02005-6         -1.586      0.250   -6.34  8.4e-10 ***
## year02006-7         -1.754      0.250   -7.01  1.6e-11 ***
## year02007-8         -1.819      0.250   -7.26  3.3e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.3 on 299 degrees of freedom
## Multiple R-squared:  0.38,   Adjusted R-squared:  0.368
## F-statistic: 30.6 on 6 and 299 DF,  p-value: <2e-16
```

Regression results from Harper et al., 2012

Regression equation

$$Y_{st} = \beta_0 + \beta_1 MML_{st} + \gamma_t + \delta_s + \varepsilon_{st}$$

- Y(st) = Marijuana use in state s in year t
- MML(st) = A dummy variable indicating whether or not a state had a MML in place in year t
- Gamma(t) = A fixed effect for each year
- Delta(a) = A fixed effect for each state
- Epsilon(st) = A state-year-specific error term.

Year fixed effects control for any secular trend affecting marijuana use that is common to all states (not constrained to be linear).

State fixed effects control for any time-invariant characteristics of states. States that passed laws are our treatment group, and we use states that did not pass laws as a control group to estimate the counterfactual trend that treatment states would have demonstrated, had we been able to observe them.

Under this specification, the effect of the law is identified by comparing within-state changes in marijuana use before and after the passage of a law in states passing laws to states whose law status does not change.

```r
DiffInDiff <- lm(pmu ~ factor(mmlf) + year0 + factor(state), data = mml_data_12to17,
    na.action = na.exclude)
summary(DiffInDiff)
```

```
##
## Call:
## lm(formula = pmu ~ factor(mmlf) + year0 + factor(state), data = mml_data_12to17,
##     na.action = na.exclude)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5749 -0.3538 -0.0177  0.3005  1.9493
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         7.0887     0.2584   27.43  < 2e-16 ***
## factor(mmlf)yes                    -0.5937     0.2689   -2.21  0.02816 *
## year02003-4                        -0.4463     0.1199   -3.72  0.00024 ***
## year02004-5                        -1.0271     0.1199   -8.57  1.1e-15 ***
## year02005-6                        -1.4410     0.1205  -11.96  < 2e-16 ***
## year02006-7                        -1.5607     0.1213  -12.87  < 2e-16 ***
## year02007-8                        -1.5771     0.1223  -12.89  < 2e-16 ***
## factor(state)Alaska                 4.0220     0.4400    9.14  < 2e-16 ***
```

```
## factor(state)Arizona                   1.2650     0.3482    3.63  0.00034 ***
## factor(state)Arkansas                  1.3117     0.3482    3.77  0.00021 ***
## factor(state)California                1.8270     0.4400    4.15  4.5e-05 ***
## factor(state)Colorado                  3.4437     0.4400    7.83  1.4e-13 ***
## factor(state)Connecticut               2.3267     0.3482    6.68  1.5e-10 ***
## factor(state)Delaware                  1.4983     0.3482    4.30  2.4e-05 ***
## factor(state)District of Columbia      0.7600     0.3482    2.18  0.03002 *
## factor(state)Florida                   1.1617     0.3482    3.34  0.00098 ***
## factor(state)Georgia                  -0.0383     0.3482   -0.11  0.91244
## factor(state)Hawaii                    2.6120     0.4400    5.94  9.7e-09 ***
## factor(state)Idaho                     0.6600     0.3482    1.90  0.05922 .
## factor(state)Illinois                  0.4833     0.3482    1.39  0.16641
## factor(state)Indiana                   1.0200     0.3482    2.93  0.00372 **
## factor(state)Iowa                      0.0200     0.3482    0.06  0.95425
## factor(state)Kansas                    0.8467     0.3482    2.43  0.01575 *
## factor(state)Kentucky                  1.3717     0.3482    3.94  0.00011 ***
## factor(state)Louisiana                -0.1200     0.3482   -0.34  0.73070
## factor(state)Maine                     5.2470     0.4400   11.93  < 2e-16 ***
## factor(state)Maryland                  0.5600     0.3482    1.61  0.10909
## factor(state)Massachusetts             3.2617     0.3482    9.37  < 2e-16 ***
## factor(state)Michigan                  1.9739     0.3511    5.62  5.1e-08 ***
## factor(state)Minnesota                 1.5950     0.3482    4.58  7.3e-06 ***
## factor(state)Mississippi              -0.7483     0.3482   -2.15  0.03261 *
## factor(state)Missouri                  0.9667     0.3482    2.78  0.00592 **
## factor(state)Montana                   4.4547     0.4141   10.76  < 2e-16 ***
## factor(state)Nebraska                  0.5333     0.3482    1.53  0.12692
## factor(state)Nevada                    2.2537     0.4400    5.12  6.0e-07 ***
## factor(state)New Hampshire             3.4167     0.3482    9.81  < 2e-16 ***
## factor(state)New Jersey                0.3450     0.3482    0.99  0.32281
## factor(state)New Mexico                3.4646     0.3596    9.63  < 2e-16 ***
## factor(state)New York                  2.0217     0.3482    5.81  1.9e-08 ***
## factor(state)North Carolina            1.2100     0.3482    3.47  0.00060 ***
## factor(state)North Dakota             -0.1233     0.3482   -0.35  0.72352
## factor(state)Ohio                      1.5967     0.3482    4.58  7.2e-06 ***
## factor(state)Oklahoma                  0.8317     0.3482    2.39  0.01768 *
## factor(state)Oregon                    3.3070     0.4400    7.52  1.0e-12 ***
## factor(state)Pennsylvania              0.9150     0.3482    2.63  0.00914 **
## factor(state)Rhode Island              4.4818     0.3733   12.01  < 2e-16 ***
## factor(state)South Carolina            0.1167     0.3482    0.34  0.73790
## factor(state)South Dakota              0.8583     0.3482    2.46  0.01439 *
## factor(state)Tennessee                 0.0183     0.3482    0.05  0.95806
## factor(state)Texas                    -0.2683     0.3482   -0.77  0.44172
## factor(state)Utah                     -1.0383     0.3482   -2.98  0.00315 **
## factor(state)Vermont                   5.5630     0.4141   13.43  < 2e-16 ***
## factor(state)Virginia                  0.8233     0.3482    2.36  0.01884 *
## factor(state)Washington                2.2087     0.4400    5.02  9.8e-07 ***
## factor(state)West Virginia             1.0567     0.3482    3.03  0.00267 **
## factor(state)Wisconsin                 1.4367     0.3482    4.13  5.0e-05 ***
## factor(state)Wyoming                   1.2350     0.3482    3.55  0.00047 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6 on 249 degrees of freedom
## Multiple R-squared:  0.882,  Adjusted R-squared:  0.856
## F-statistic: 33.3 on 56 and 249 DF,  p-value: <2e-16
```

Additional notes:

An in depth understanding of the model specification is critical.

- There are lots of really good examples out there in the literature. Mostly these are in economics and some in epidemiology.

Many potential applications in active living research:

- Fuller D, Gauvin L, Kestens Y, et al. Impact Evaluation of a Public Bicycle Share Program on Cycling: A

Case Example of BIXI in Montreal, Quebec. American Journal of Public Health. 2013;103(3):e1â€“e8.

The end