

# DID WITH POOLED CROSS-SECTIONS

Ozan Bakış<sup>1</sup>

<sup>1</sup>Bahcesehir University, Department of Economics and BETAM

# Outline

---

## 1 Regression with pooled cross-sections

- Basics

- DD estimator using Pooled CS

## Pooled cross section: overview I

---

- Data obtained by pooling cross sections (PCS) are very useful for establishing trends and conducting policy analysis.
- A pooled cross section is available whenever a survey is repeated over time with new random samples obtained in each time period.
- Examples include the Current Population Survey (CPS) in USA and Household Labor Survey (Hanehalkı İşgücü Anketi) in Turkey.
- With a PCS, often a goal is to see how the mean value of a variable (fertility) has changed over time in ways that cannot be explained by observable variables (education).
- **Ex.:** Has the fertility rate changed in ways that cannot be explained by education?

## Pooled cross section: overview II

---

- From a policy perspective, PCSs are at the foundation of difference-in-differences estimation.
- The typical DD setup is that data can be collected both before and after an intervention (or "treatment"), and there is (at least) one "control group" and (at least) one "treatment" group.
- Often the intervention is of a yes/no form. But other nonbinary treatments (such as class size) can be handled, too.

# Application I

---

- IS the Change in women's fertility in the USA (1972-1984) can be explained by rise in education levels of women?
- How much of the fall in average fertility cannot be explained by changes in observed factors, including education? Here we require a PCS and look at coefficients on year dummies.
- How much of the overall fall in average fertility be explained by increases in average education?
- Before a full regression model let us go step by step and try to understand the underlying patterns. What is the trend for average number of kids over years?

```
f_url = "https://github.com/obakis/econ_data/raw/master/fertil1.rds"  
download.file(url = f_url, destfile = "fertil1.rds", mode="wb")  
dat = readRDS("fertil1.rds")  
library(lmtest) # for pretty print of reg results
```

## Application II

---

```
#with(dat, tapply(kids, year, FUN=summary))
aggregate(kids ~ year, FUN=mean, data=dat)

##   year kids
## 1    72 3.03
## 2    74 3.21
## 3    76 2.80
## 4    78 2.80
## 5    80 2.82
## 6    82 2.40
## 7    84 2.24
```

- The average fertility rate fell by about 0.79,  $-0.79 = 2.24 - 3.03$ .
- The same could be done through a regression as well.

```
reg1 = lm(kids ~year, data=dat) # this is probably not what we want
coeftest(reg1) # or maybe it is???
```

## Application III

---

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.4256     0.9267   9.09 < 2e-16 ***
## year         -0.0727     0.0118  -6.14 1.1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

dat$year_f = factor(dat$year)
reg2 = lm(kids ~ year_f, data=dat) # this is probably a better way
coefest(reg2) # You see a trend over time?

##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.026     0.130  23.24 < 2e-16 ***
## year_f74         0.182     0.180   1.02 0.30976
## year_f76        -0.223     0.185  -1.20 0.22912
## year_f78        -0.221     0.188  -1.18 0.23975
## year_f80        -0.209     0.189  -1.11 0.26865
```

## Application IV

---

```
## year_f82      -0.622      0.177    -3.53  0.00044 ***
## year_f84      -0.788      0.179    -4.41  1.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Again we can say that the average fertility rate fell by about **0.79**. But more importantly we have an idea about its significance. This is the benefit of regression compared to simple comparison of means!
- Education is an important determinant of fertility. Let us see how it changes over years

```
aggregate(educ ~ year, FUN=mean, data=dat)
```

```
##   year educ
## 1   72 12.2
## 2   74 12.3
## 3   76 12.2
## 4   78 12.6
## 5   80 12.9
## 6   82 13.2
## 7   84 13.3
```

# Application V

---

```
reg3 = lm(educ ~ year_f, data=dat) # this is probably OK
coef(summary(reg3)) # You see a trend over time?
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.1538      0.209  58.164 0.000000
## year_f74     0.1467      0.288   0.509 0.610709
## year_f76     0.0764      0.297   0.257 0.797296
## year_f78     0.4895      0.302   1.620 0.105495
## year_f80     0.7264      0.303   2.400 0.016566
## year_f82     1.0720      0.283   3.783 0.000163
## year_f84     1.1117      0.287   3.879 0.000111
```

- Overall the increase in mean education is 1.11 years. To see the effect of this increase on fertility we need to know the partial effect of education on fertility. For this we run the following regression

```
dat$year_f = factor(dat$year)
reg4 <- lm(kids~educ+age+agesq+black+east+northcen+west+year_f, data=dat)
coef(summary(reg4))
```

## Application VI

---

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	-7.95228	3.05004	-2.607	9.25e-03
## educ	-0.12269	0.01803	-6.806	1.64e-11
## age	0.53904	0.13837	3.896	1.04e-04
## agesq	-0.00588	0.00156	-3.762	1.78e-04
## black	1.09095	0.17311	6.302	4.22e-10
## east	0.25290	0.12685	1.994	4.64e-02
## northcen	0.38523	0.11853	3.250	1.19e-03
## west	0.23257	0.16532	1.407	1.60e-01
## year_f74	0.25608	0.17265	1.483	1.38e-01
## year_f76	-0.10630	0.17857	-0.595	5.52e-01
## year_f78	-0.07050	0.18136	-0.389	6.98e-01
## year_f80	-0.07855	0.18261	-0.430	6.67e-01
## year_f82	-0.53255	0.17230	-3.091	2.05e-03
## year_f84	-0.54226	0.17436	-3.110	1.92e-03

- Each additional year of education is estimated to reduce the number of children by about **0.123**, on average.

## Application VII

---

- Compared to 1972, fertility fell by about 0.55 children in 1984. This is the drop that cannot be explained by the explanatory variables.
- Of the overall drop of about 0.79 children, the increase in education (1.11 years on average) accounts for about  $0.14 \approx 1.11 \times 0.123$  of that, or about 18%.
- In the previous estimation with the fertility data, we assumed the effect of education (and all other variables) was the same over time.
- We can easily allow the slopes to change over time by forming interactions and adding them to the model.

```
reg5 = lm(kids~educ+age+agesq+black+east+northcen+west+
          year_f+year_f:educ, data=dat)
coef(summary(reg5))
```

## Application VIII

---

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	-8.61295	3.12654	-2.755	5.97e-03
## educ	-0.02456	0.05357	-0.458	6.47e-01
## age	0.51493	0.13894	3.706	2.21e-04
## agesq	-0.00561	0.00157	-3.573	3.67e-04
## black	1.09031	0.17333	6.290	4.55e-10
## east	0.24805	0.12711	1.952	5.12e-02
## northcen	0.37446	0.11869	3.155	1.65e-03
## west	0.21657	0.16582	1.306	1.92e-01
## year_f74	0.82860	0.90059	0.920	3.58e-01
## year_f76	0.89183	0.87799	1.016	3.10e-01
## year_f78	1.71434	0.94992	1.805	7.14e-02
## year_f80	0.97952	0.89447	1.095	2.74e-01
## year_f82	1.06195	0.87204	1.218	2.24e-01
## year_f84	1.54243	0.89484	1.724	8.50e-02
## educ:year_f74	-0.04779	0.07230	-0.661	5.09e-01
## educ:year_f76	-0.08241	0.07058	-1.168	2.43e-01
## educ:year_f78	-0.14509	0.07515	-1.931	5.38e-02
## educ:year_f80	-0.08768	0.07024	-1.248	2.12e-01
## educ:year_f82	-0.12869	0.06807	-1.891	5.89e-02
## educ:year_f84	-0.16553	0.06961	-2.378	1.76e-02

# Application IX

---

Let us test year and education interaction terms:

```
library(car)
```

```
linearHypothesis(reg5, matchCoefs(reg5, "educ:year"), vcov=hccm(reg5))
```

```
## Linear hypothesis test
```

```
##
```

```
## Hypothesis:
```

```
## educ:year_f74 = 0
```

```
## educ:year_f76 = 0
```

```
## educ:year_f78 = 0
```

```
## educ:year_f80 = 0
```

```
## educ:year_f82 = 0
```

```
## educ:year_f84 = 0
```

```
##
```

```
## Model 1: restricted model
```

```
## Model 2: kids ~ educ + age + agesq + black + east + northcen + west +
```

```
##   year_f + year_f:educ
```

```
##
```

```
## Note: Coefficient covariance matrix supplied.
```

```
##
```

```
##   Res.Df Df    F Pr(>F)
```

## Application X

---

```
## 1 1115
## 2 1109 6 1.02 0.41
```

Jointly insignificant, even though `educ:year_f84` and `educ:year_f82` are individually significant.

- Coefficient on, say, `year_f84` is the difference in fertility between 1984 and 1972 at `educ = 0`; not interesting.
- Effect of schooling in base year very close to zero.
- The joint test for all interactions with `educ` gives `p-value = 0.41`, so we cannot reject the null that the effect of education has been constant. But it seems fertility has become more sensitive to education in the last couple of years of the data (1982, 1984).

## DD with 2 groups and 2 time periods I

- Useful to study the data coming from a **natural experiment** (or a quasi-experiment). This is called natural experiment because
  - ⇒ an exogenous event (usually a change in government policy) changes the conditions under which individuals / firms etc. operate.
  - ⇒ There are at least one control group (not affected by policy change) and one treatment group (affected by policy change).
- DD methodology is used widely to evaluate the consequences of **natural experiments** (or quasi-experiments). There are two key elements in natural experiments:
- Outcomes are observed for two groups over two time periods. One of the groups is exposed to a "treatment" (or intervention) in the second period but not in the first period. The second group is not exposed to the treatment during either period.

## DD with 2 groups and 2 time periods II

- In the textbook version where we have 2 groups and 2 time periods. As a result, treating one of groups and time periods as reference, we need at least two dummy variables: 1 for the other group and 1 for the other time period.
- **Ex.:** Remember the famous minimum wage example: Card and Krueger (1994) studied the increase in the minimum wage in New Jersey from 4.25 USD to 5.05 USD. This change took effect on April 1, 1992. The minimum wage in Pennsylvania remained at 4.25 USD throughout this period.
- Card and Krueger collected data on employment at fast food restaurants in New Jersey and Pennsylvania (the neighboring state) in February (before) and in November (after).
- Here New Jersey is the "treatment state" and Pennsylvania is the "control state". Without data from Pennsylvania, we cannot control for aggregate changes over time that affect employment in both states.

## DD with 2 groups and 2 time periods III

- Let **A** be the control group and **B** the treatment group. Let  $dT = 1$  for a unit in B, and  $dT = 0$  for a unit in A in both periods. Similarly, let  $d2 = 1$  for a unit in the second time period and  $d2 = 0$  for a unit in the first period for both states. A typical model would be:

$$y = \beta_0 + \beta_1 dT + \delta_0 d2 + \delta_1 d2 \cdot dT + u \quad (1)$$

where  $y$  is the outcome of interest. What was missing in previous regression is the specificity of the second period for treatment group !

	Before (1)	After (2)	After – Before
Control (A)	$\beta_0$	$\beta_0 + \delta_0$	$\delta_0$
Treatment (B)	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
Treatment – Control	$\beta_1$	$\beta_1 + \delta_1$	$\delta_1$

- While powerful, the basic DD approach can suffer from several problems.

## DD with 2 groups and 2 time periods IV

---

- First, there may be compositional effects. For example, when studying the effects of the intervention to reduce class size, it may be the case that the students in the two years are not comparable. Perhaps the smaller class sizes attracts new students to that district. This is the problem of compositional changes.  
⇒ One can control for changes in composition to some extent by including observed covariates as controls.

$$y = \beta_0 + \beta_1 dT + \delta_0 d2 + \delta_1 d2 \cdot dT + \mathbf{x}\gamma + u$$

- Second, a potential problem with using only two periods is that the control and treatment groups may be trending at different rates having nothing to do with the intervention. There is a large literature on dealing with violation of parallel trends assumption.  
⇒ Only way to solve this problem is get another control group or more years of data.

# Application I

---

## Effect of Direct Sale Points ("tanzim satış") on inflation rate in Turkey

- On 11 February 2019, the government introduced Direct Sale Points (DSP). DSPs are municipality-led vegetable stalls which sell products at discounted prices (called "tanzim satış" in Turkish).
- Only 9 products have been sold at these DSP (tomatoes, potatoes, onions, green peppers and other vegetables). The data contains monthly inflation rate in February for these 9 products in all regions of Turkey in 2018 (before DSP) and in 2019 (after DSP).
- In February 2019 DSP were operated only in Istanbul and Ankara (called "tanzim regions"), later they were operated in other cities as well. Since there were only 65 DSP in Tanzim regions (50 in Istanbul and 15 in Ankara) but a large population (15 million in Istanbul and almost 6 million in Ankara) it is not obvious that there is an effect on inflation rates.

## Application II

---

```
furl = "https://github.com/obakis/econ_data/raw/master/tanzim_pooled.rds"  
download.file(furl, destfile = "tanzim_pooled.rds", mode="wb")  
tanzim = readRDS("tanzim_pooled.rds")
```

- A naive analyst would use data on tanzim regions and estimate a simple model where the year dummy on 2019 (y19) capture the effect of DSPs on inflation rate:

$$\text{inf} = \alpha_0 + \alpha_1 y19 + u$$

```
reg1 = lm(inf ~ y19, data=subset(tanzim, tanzim_reg==1))  
coef(summary(reg1))
```

```
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept)      5.67         2.93   1.94  0.06114  
## y19             -12.11         4.14  -2.92  0.00614
```

## Application III

---

- Since this is a simple regression on a single dummy variable, the intercept is the average inflation rate in 2018 and the coefficient on  $y19r$  is the difference in the inflation rate from 2018 to 2019 in tanzim regions. Since one thing that changes is the introduction of DSPs, a naive approach would be to conclude that the effect of DSPs on inflation rate is  $-12.1$  percentage points.
- Unfortunately,  $\hat{\alpha}_1$  will have a causal interpretation only when Gauss-Markov assumptions are verified. Among them the most important one is zero-conditional mean:  $Cov(y19, u) = 0$ . For this to be true, we need that nothing in the error term changes from 2018 to 2019 except.
- It is highly possible that some other things are changing over time along with the introduction of DSPs. Macro shocks are such an example. Indeed, if we run the same regression for other regions, we obtain

```
reg2 = lm(inf ~ y19, data=subset(tanzim, tanzim_reg==0))  
coef(summary(reg2))
```

## Application IV

---

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	6.59	0.81	8.14	4.33e-15
## y19	-5.65	1.15	-4.93	1.17e-06

- Therefore, even in other regions without DSPs the inflation rate was lower in 2019 which supports the claim about macro shocks specific to 2019.
- If the estimate we get from the second regression using data from other regions can be thought as a proxy for macro shocks (or all other effects other than DSPs), then what would be the "true" effect of DSPS? Probably,  $-12.1 - (-5.65) = -6.45$ . This is the answer to the question: how 2019 is different in tanzim regions compared to other regions?

## Application V

---

- Another naive analyst could use data from 2019 and estimate a simple model where the coefficient on dummy variable `tanzim_reg` capture the effect of DSPs on inflation rate:

$$\text{inf} = \gamma_0 + \gamma_1 \text{tanzim\_reg} + u$$

```
reg1 = lm(inf ~ tanzim_reg, data=subset(tanzim, year==2019))  
coef(summary(reg1))
```

```
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept)    0.943      0.67    1.41  0.16076  
## tanzim_reg    -7.376      2.42   -3.05  0.00254
```

- The intercept is the average inflation rate in other regions and the coefficient on `tanzim_reg` is the difference in the inflation rate between `tanzim` regions and the rest of Turkey.

## Application VI

---

- Unfortunately,  $\hat{\gamma}_1$  will have a causal interpretation only when  $\text{Cov}(\text{tanzim\_reg}, \mathbf{u}) = 0$ . For this to be true, we need that nothing in the error term changes between tanzim and other regions.
- But in reality there may be region-specific unobserved factors in the error term. changes from 2018 to 2019 except. Indeed, if we run the same regression using data from 2018, we obtain

```
reg2 = lm(inf ~ tanzim_reg, data=subset(tanzim, year==2018))
coef(summary(reg2))
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.591      0.933   7.064 1.87e-11
## tanzim_reg    -0.919      3.365  -0.273 7.85e-01
```

- Therefore, even before the introduction of DSPs the inflation rate was lower in tanzim regions.

## Application VII

---

- If the estimate for region-specific unobserved factors is given by the second regression, then what would be the "true" effect of DSPS? Probably,  $-7.38 - (-0.92) = -6.46$ . This is the answer to the question: how tanzim regions are different from other regions in 2019 compared to 2018 ?
- Both of above the reasonings come to see how the difference between tanzim regions and other regions change from 2018 to 2019. This is known as the difference-in-differences (DD) estimator because it can be expressed as

$$\hat{\delta}_1 = (\bar{\text{inf}}_{19,t} - \bar{\text{inf}}_{18,t}) - (\bar{\text{inf}}_{19,o} - \bar{\text{inf}}_{18,o})$$

or

$$\hat{\delta}_1 = (\bar{\text{inf}}_{19,t} - \bar{\text{inf}}_{19,o}) - (\bar{\text{inf}}_{18,t} - \bar{\text{inf}}_{18,o})$$

where **o** means "other regions" and **t** means "tanzim regions". To calculate the effect of DSP using DD estimator

```
aggregate(inf ~ year + tanzim_reg, data=tanzim, FUN=mean)
```

## Application VIII

---

```
##   year tanzim_reg   inf
## 1 2018           0 6.591
## 2 2019           0 0.943
## 3 2018           1 5.672
## 4 2019           1 -6.433

dT = -6.43-5.67
dT
## [1] -12.1

dC = 0.94 - 6.59
dC
## [1] -5.65

dT - dC
## [1] -6.45
```

- An important question is whether  $\hat{\delta}_1$  is different from zero? For this, we need std. error of it. How can we compute a std. error for  $\hat{\delta}_1$ ?

## Application IX

- Actually this can be obtained through the following model

$$\text{inf} = \beta_0 + \delta_0 y19 + \beta_1 \text{tanzim\_reg} + \delta_1 \text{tanzim\_reg} \cdot y19 + u$$

The following table summarizes the meaning of various coefficients:

$\bar{y}$ is equal to	Before	After	After – Before
Control (other)	$\beta_0$	$\beta_0 + \delta_0$	$\delta_0$
Treatment (tanzim)	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
Treatment – Control			$\delta_1$

- $\beta_0$ : average inflation in control group (other regions) in the first period.
- $\beta_1$ : difference between treatment (tanzim) and control (other) groups in the first period.
- $\delta_0$ : change in inflation rate across the two periods for the control group.
- $\beta_1 + \delta_1$ : difference between treatment and control groups after the intervention (DSPs).

# Application X

---

- $\delta_0 + \delta_1$ : change in outcome across the two periods for the treatment group.
- $\delta_1$  is the DD estimate.
- In R:

```
reg = lm(inf ~ y19*tanzim_reg, data=tanzim)
summary(reg)

##
## Call:
## lm(formula = inf ~ y19 * tanzim_reg, data = tanzim)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.58  -9.63  -1.90   8.25  40.02
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.591      0.813   8.11  4.5e-15 ***
## y19              -5.648      1.149  -4.92  1.2e-06 ***
## tanzim_reg       -0.919      2.930  -0.31   0.75
```

# Application XI

---

```
## y19:tanzim_reg   -6.457      4.143   -1.56    0.12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.9 on 464 degrees of freedom
## Multiple R-squared:  0.0746, Adjusted R-squared:  0.0686
## F-statistic: 12.5 on 3 and 464 DF,  p-value: 7.4e-08
### A better model would allow for clustered std errors
library("fixest")
reg2 = feols(inf ~ y19*tanzim_reg, data=tanzim, cluster="nuts2")
summary(reg2)

## OLS estimation, Dep. Var.: inf
## Observations: 468
## Standard-errors: Clustered (nuts2)
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)      6.591      0.539   12.23 4.7672e-12 ***
## y19              -5.648      0.943   -5.99 2.9829e-06 ***
## tanzim_reg       -0.919      0.861   -1.07 2.9581e-01
## y19:tanzim_reg   -6.457      1.383   -4.67 8.8010e-05 ***
## ---
```

## Application XII

---

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 11.9   Adj. R2: 0.068642

### Alternatively, we can avoid nuts2 and year fixed effects being printed.
#reg2 = feols(inf ~ tanzim_reg:y19 | nuts2+year,
#             data=tanzim, cluster="nuts2")
```

# When treatment is continuous I

---

## Effect of a Garbage Incinerator's Location on Housing Prices

- Above, we mentioned problem of compositional changes. This example will allow us to see how to control for compositional changes in the DD approach.
- We study the effect that a new garbage incinerator had on housing values in Massachusetts. The rumor that a new incinerator would be built began after 1978, and construction began in 1981.
- The incinerator was expected to be in operation soon after the start of construction. We will use data on prices of houses that sold in 1978 and another sample on those that sold in 1981.
- Year dummy is  $y_{81}$  it takes the value of 1 for year 1981 and 0 for year 1978. We also define  $near_{inc} = 1$  if a house is near the incinerator, more concertely if it is within three miles.

## When treatment is continuous II

---

```
f_url = "https://github.com/obakis/econ_data/raw/master/kielmc.rds"
download.file(url = f_url, destfile = "kielmc.rds", mode="wb")
dat = readRDS("kielmc.rds")
```

- DD estimator using sample averages

```
aggregate(rprice~year, FUN=mean, data=dat)
```

```
##   year rprice
## 1 1978  76628
## 2 1981  92663
```

- DD estimator using regression

$$rprice = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 nearinc \cdot y81 + u$$

- The parameter of interest is,  $\delta_1$ , which measures the decline in housing values due to the new incinerator, provided we assume that houses both near and far from the site did not appreciate at different rates for other reasons.

## When treatment is continuous III

---

```
#did <- lm(rprice ~ y81 + nearinc + nearinc:y81, data=kielmc)
did <- lm(rprice ~ nearinc*y81, data=dat)
coef(summary(did))

##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    82517      2727   30.26 1.71e-95
## nearinc       -18824      4875   -3.86 1.37e-04
## y81           18790      4050    4.64 5.12e-06
## nearinc:y81   -11864      7457   -1.59 1.13e-01
```

- To control for compositional changes we will add other control variables that control for the change in average house attributes (quality, space etc.) over time which may depend on the location of houses.
- Even if the composition of houses does not change that much, including house characteristics will reduce the error variance, which shrinks the standard error of the OLS estimates.

## When treatment is continuous IV

---

```
#intst: dist. to interstate, ft
#bath: number of bathrooms
#rooms: number of rooms
did_ctr <- lm(log(rprice) ~ nearinc*y81 + age + log(intst)
              + rooms + baths, data=dat)
coef(summary(did_ctr))
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	9.57744	0.284799	33.629	6.20e-106
## nearinc	0.01687	0.052886	0.319	7.50e-01
## y81	0.17751	0.031309	5.670	3.25e-08
## age	-0.00188	0.000466	-4.030	7.01e-05
## log(intst)	0.05043	0.027763	1.816	7.03e-02
## rooms	0.08714	0.018847	4.623	5.53e-06
## baths	0.25285	0.023491	10.763	3.40e-23
## nearinc:y81	-0.07528	0.057386	-1.312	1.91e-01

Comparing both:

## When treatment is continuous V

	Dependent variable:	
	rprice (1)	log(rprice) (2)
nearinc	-18,824.000*** (4,875.000)	0.017 (0.053)
y81	18,790.000*** (4,050.000)	0.178*** (0.031)
age		-0.002*** (0.0005)
log(intst)		0.050* (0.028)
rooms		0.087*** (0.019)
baths		0.253*** (0.023)
nearinc:y81	-11,864.000 (7,457.000)	-0.075 (0.057)
Constant	82,517.000*** (2,727.000)	9.580*** (0.285)
Observations	321	321
Adjusted R <sup>2</sup>	0.166	0.650

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01