

Causal Inference - Problem Set 3

Gustavo Baroni

9/5/2020

Question 1

Explain how we can make use of stratification to solve the causal inference problem.

We can make use of stratification to solve the causal inference problem because doing so enables one to concatenate individuals with common characteristics in subgroups. Therefore, stratification can be used by any researcher to make comparisons between well defined groups.

Question 2

Consider the assumption: $D_i \perp (Y_i(1), Y_i(0)) | X_i$ ¹. Interpret the assumption. Is this often true in most empirical applications (give examples)? Is it testable?

The assumption above means that the outputs for $Y_i(1)$ and $Y_i(0)$ are given no matter whether i was treated. In other words, the realizations are independent of the treatment or control effect. Donald Rubin developed the *Stable Unit-Treatment-Value Assumption (SUTVA)* in which the assumption made in this exercise is present. He presented an example about the aspirin treatment effect in headaches. Rubin argue that, if two people are in the same experiment together, chances are that someone (in the control group, for example) complain about the headache to the partner (in the treatment group). That way, the one in the treatment group could still have some headache due to the complain. Hence, the treatment for one individual would effect the another output. It is important to affirm that, despite of Rubin had cooked the data and just calculated the difference in means for each group, the assumption behind the SUTVA is crucial to get a more accurate treatment effect from a experiment.

Question 3

After constructing stratas, we can make use of blocking estimators to estimate average treatment effects. Present the simple blocking estimator.

A simple blocking estimator is the one for control sexes. In a experiment, a researcher could use the biological condition of being *male* and *female* as a blocking estimator in order to reduce the variance and the variability. Doing so, the researcher would lead the experiment to grater precision.

Question 4

Explain the differences between OLS and the matching estimands.

As *pointed out Jorn-Steffen Pischke*, the main difference between Regression and Matching is the weight they use to combine the covariate specific treatment effects. Matching technique uses $P(D_i = 1 | X_i = x)$ and is better used where there are many treated observations. Regression, on the other hand, uses $P(D_i = 1 | X_i = x)(1 - P(D_i = 1 | X_i = x))$ and is better used where there are half treated and half untreated observations because the calls will produce the lowest variance.

¹Unfortunately, I could not do independent symbol with a *Latex command*.

Question 5

This exercise aims to assess the impact of a training program (National Supported Work Demonstration, NSW from now on) on earnings. NSW was a temporary employment program designed to help disadvantaged workers lacking basic job skills move into labor market by giving them work experience and counseling in a sheltered environment.

The program was run as a field experiment, where participants were randomly assigned to treatment and control groups. The random assignment was explored by Lalonde (1986)², in order to estimate its causal effect on wages.

To discuss causal inference and sample selection bias in nonexperimental settings, Dehejia and Wahba (1999, 2002)³ combine the treated units from a randomized evaluation of the NSW with nonexperimental comparison units drawn from survey datasets (CPS and PSID) to form a composite dataset. Dehejia and Wahba (1999, 2002) thus compare the estimates of the treatment effect obtained using propensity score matching on the composite dataset to the benchmark results from the field experiment.

Using the experimental dataset `nsw.dta`, compute the following:

- (a) Proceed a balance check on pre-treatment variables (i.e. for each pre-treatment variable, compute its mean and standard error for treatment and control groups and test if the difference in means between groups is statistically significant).

Proceeding the *balance checks*:

```
library(Matching)
library(dplyr)

treat <- filter (nsw, treat == 1)
no_treat <- filter (nsw, treat == 0)

## Testing the balance for age
age <- balanceUV(treat$age, no_treat$age)
summary(age)

## mean treatment..... 25.816
## mean control..... 25.054
## std mean diff..... 10.655
##
## mean raw eQQ diff..... 0.94054
## med raw eQQ diff..... 1
## max raw eQQ diff..... 7
##
## mean eCDF diff..... 0.025364
## med eCDF diff..... 0.022193
## max eCDF diff..... 0.065177
##
## var ratio (Tr/Co)..... 1.0278
## T-test p-value..... 0.26594

## Testing the balance for education
educ <- balanceUV(treat$educ, no_treat$educ)
summary(educ)
```

²Lalonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. The American Economic Review, 604-620.

³Dehejia, R. H., Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. Journal of the American statistical Association, 94(448), 1053-1062. Dehejia, R. H., Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. Review of Economics and statistics, 84(1), 151-161.

```

## mean treatment..... 10.346
## mean control..... 10.088
## std mean diff..... 12.806
##
## mean raw eQQ diff..... 0.40541
## med  raw eQQ diff..... 0
## max  raw eQQ diff..... 2
##
## mean eCDF diff..... 0.028698
## med  eCDF diff..... 0.012682
## max  eCDF diff..... 0.12651
##
## var ratio (Tr/Co)..... 1.5513
## T-test p-value..... 0.15017

## Testing the balance for phenotype
black <- balanceUV(treat$black, no_treat$black)
summary(black)

## mean treatment..... 0.84324
## mean control..... 0.82692
## std mean diff..... 4.4767
##
## mean raw eQQ diff..... 0.016216
## med  raw eQQ diff..... 0
## max  raw eQQ diff..... 1
##
## mean eCDF diff..... 0.0081601
## med  eCDF diff..... 0.0081601
## max  eCDF diff..... 0.01632
##
## var ratio (Tr/Co)..... 0.92503
## T-test p-value..... 0.64736

## Testing the balance for ethnicity
hisp <- balanceUV(treat$hisp, no_treat$hisp)
summary(hisp)

## mean treatment..... 0.059459
## mean control..... 0.10769
## std mean diff..... -20.341
##
## mean raw eQQ diff..... 0.048649
## med  raw eQQ diff..... 0
## max  raw eQQ diff..... 1
##
## mean eCDF diff..... 0.024116
## med  eCDF diff..... 0.024116
## max  eCDF diff..... 0.048233
##
## var ratio (Tr/Co)..... 0.58288
## T-test p-value..... 0.064043

## Testing the balance for marital status
married <- balanceUV(treat$married, no_treat$married)
summary(married)

```

```
## mean treatment..... 0.18919
## mean control..... 0.15385
## std mean diff..... 8.9995
##
## mean raw eQQ diff..... 0.037838
## med raw eQQ diff..... 0
## max raw eQQ diff..... 1
##
## mean eCDF diff..... 0.017672
## med eCDF diff..... 0.017672
## max eCDF diff..... 0.035343
##
## var ratio (Tr/Co)..... 1.1802
## T-test p-value..... 0.33425
## Testing the balance for earnings in 1974
re74 <- balanceUV(treat$re74, no_treat$re74)
summary(re74)
```

```
## mean treatment..... 2095.6
## mean control..... 2107
## std mean diff..... -0.23437
##
## mean raw eQQ diff..... 487.98
## med raw eQQ diff..... 0
## max raw eQQ diff..... 8413
##
## mean eCDF diff..... 0.019223
## med eCDF diff..... 0.0158
## max eCDF diff..... 0.047089
##
## var ratio (Tr/Co)..... 0.7381
## T-test p-value..... 0.98186
```

```
## Testing the balance for earnings in 1975
re75 <- balanceUV(treat$re75, no_treat$re75)
summary(re75)
```

```
## mean treatment..... 1532.1
## mean control..... 1266.9
## std mean diff..... 8.2363
##
## mean raw eQQ diff..... 367.61
## med raw eQQ diff..... 0
## max raw eQQ diff..... 2110.2
##
## mean eCDF diff..... 0.050834
## med eCDF diff..... 0.061954
## max eCDF diff..... 0.10748
##
## var ratio (Tr/Co)..... 1.0763
## T-test p-value..... 0.38527
```

```
## Testing the balance for unemployment status in 1974
u74 <- balanceUV(treat$u74, no_treat$u74)
summary(u74)
```

```
## mean treatment..... 0.70811
## mean control..... 0.75
## std mean diff..... -9.1895
##
## mean raw eQQ diff..... 0.037838
## med raw eQQ diff..... 0
## max raw eQQ diff..... 1
##
## mean eCDF diff..... 0.020946
## med eCDF diff..... 0.020946
## max eCDF diff..... 0.041892
##
## var ratio (Tr/Co)..... 1.1041
## T-test p-value..... 0.33033
## Testing the balance for unemployment status in 1975
u75 <- balanceUV(treat$u75, no_treat$u75)
summary(u75)
```

```
## mean treatment..... 0.6
## mean control..... 0.68462
## std mean diff..... -17.225
##
## mean raw eQQ diff..... 0.081081
## med raw eQQ diff..... 0
## max raw eQQ diff..... 1
##
## mean eCDF diff..... 0.042308
## med eCDF diff..... 0.042308
## max eCDF diff..... 0.084615
##
## var ratio (Tr/Co)..... 1.1133
## T-test p-value..... 0.068031
```

As we can see, there is almost twice more black people in the control group than in the treatment. Also, Married people are a bit more present in the control group. Besides that, the treatment was fair enough balanced.

- (b) Estimate the average treatment effect of the training program (without covariates). Is it statistically different from zero? Interpret the results.

```
ate_b <- lm(re78 ~ treat, data = nsw)
summary(ate_b)

##
## Call:
## lm(formula = re78 ~ treat, data = nsw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6349  -4555  -1829   2917  53959
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4554.8      408.0  11.162 < 2e-16 ***
## treat         1794.3      632.9   2.835  0.00479 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6580 on 443 degrees of freedom
## Multiple R-squared:  0.01782,    Adjusted R-squared:  0.01561
## F-statistic: 8.039 on 1 and 443 DF,  p-value: 0.004788
```

As we can see, the treatment has an impact on the wage. Someone without the treatment would have, on average, 4554.8 and someone after the treatment would have, on average, 6349.15.

- (c) Estimate the average treatment effect of the training program, including pre-treatment covariates as control variables. Did the results change?

```
ate_c <- lm(re78 ~ treat + age + educ + black + hisp + married + re74 + re75 + u74 + u75, data = nsw)
summary(ate_c)
```

```
##
## Call:
## lm(formula = re78 ~ treat + age + educ + black + hisp + married +
##      re74 + re75 + u74 + u75, data = nsw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9610   -4353   -1566    3056   53113
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.214e+02  2.633e+03   0.084  0.93302
## treat        1.672e+03  6.343e+02   2.636  0.00869 **
## age          5.367e+01  4.530e+01   1.185  0.23681
## educ         4.029e+02  1.774e+02   2.271  0.02363 *
## black       -2.039e+03  1.164e+03  -1.752  0.08041 .
## hisp         4.246e+02  1.561e+03   0.272  0.78570
## married     -1.467e+02  8.810e+02  -0.166  0.86786
## re74         1.236e-01  8.698e-02   1.421  0.15612
## re75         1.946e-02  1.489e-01   0.131  0.89610
## u74          1.381e+03  1.186e+03   1.165  0.24474
## u75         -1.072e+03  1.023e+03  -1.048  0.29531
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6509 on 434 degrees of freedom
## Multiple R-squared:  0.05822,    Adjusted R-squared:  0.03652
## F-statistic: 2.683 on 10 and 434 DF,  p-value: 0.003392
```

The new results show that the treatment is still statistically significant (education also). Besides that, the covariates are not that significant (e.g., if we consider a married person (by almost 90% of interval), he/she would have the money he/she make dropped).

Now using the composite dataset psid.dta, compute the following:

- (d) Proceed a balance check on pre-treatment variables and discuss the results.

```
treat <- filter(psid, treat == 1)
no_treat <- filter(psid, treat == 0)

## Testing the balance for age
```

```
age <- balanceUV(treat$age, no_treat$age)
summary(age)
```

```
## mean treatment..... 25.816
## mean control..... 34.851
## std mean diff..... -126.27
##
## mean raw eQQ diff..... 9.0432
## med raw eQQ diff..... 8
## max raw eQQ diff..... 17
##
## mean eCDF diff..... 0.23165
## med eCDF diff..... 0.25299
## max eCDF diff..... 0.37714
##
## var ratio (Tr/Co)..... 0.46963
## T-test p-value..... < 2.22e-16
```

```
## Testing the balance for education
educ <- balanceUV(treat$educ, no_treat$educ)
summary(educ)
```

```
## mean treatment..... 10.346
## mean control..... 12.117
## std mean diff..... -88.077
##
## mean raw eQQ diff..... 1.8595
## med raw eQQ diff..... 2
## max raw eQQ diff..... 5
##
## mean eCDF diff..... 0.1091
## med eCDF diff..... 0.01944
## max eCDF diff..... 0.40289
##
## var ratio (Tr/Co)..... 0.42549
## T-test p-value..... < 2.22e-16
```

```
## Testing the balance for phenotype
black <- balanceUV(treat$black, no_treat$black)
summary(black)
```

```
## mean treatment..... 0.84324
## mean control..... 0.2506
## std mean diff..... 162.56
##
## mean raw eQQ diff..... 0.58919
## med raw eQQ diff..... 1
## max raw eQQ diff..... 1
##
## mean eCDF diff..... 0.29632
## med eCDF diff..... 0.29632
## max eCDF diff..... 0.59264
##
## var ratio (Tr/Co)..... 0.70739
## T-test p-value..... < 2.22e-16
```

```
## Testing the balance for ethnicity
```

```
hisp <- balanceUV(treat$hisp, no_treat$hisp)
summary(hisp)
```

```
## mean treatment..... 0.059459
## mean control..... 0.03253
## std mean diff..... 11.357
##
## mean raw eQQ diff..... 0.027027
## med raw eQQ diff..... 0
## max raw eQQ diff..... 1
##
## mean eCDF diff..... 0.013465
## med eCDF diff..... 0.013465
## max eCDF diff..... 0.026929
##
## var ratio (Tr/Co)..... 1.7859
## T-test p-value..... 0.13173
```

```
## Testing the balance for marital status
```

```
married <- balanceUV(treat$married, no_treat$married)
summary(married)
```

```
## mean treatment..... 0.18919
## mean control..... 0.86627
## std mean diff..... -172.41
##
## mean raw eQQ diff..... 0.67568
## med raw eQQ diff..... 1
## max raw eQQ diff..... 1
##
## mean eCDF diff..... 0.33854
## med eCDF diff..... 0.33854
## max eCDF diff..... 0.67708
##
## var ratio (Tr/Co)..... 1.3308
## T-test p-value..... < 2.22e-16
```

```
## Testing the balance for earnings in 1974
```

```
re74 <- balanceUV(treat$re74, no_treat$re74)
summary(re74)
```

```
## mean treatment..... 2095.6
## mean control..... 19429
## std mean diff..... -354.71
##
## mean raw eQQ diff..... 17663
## med raw eQQ diff..... 18417
## max raw eQQ diff..... 102109
##
## mean eCDF diff..... 0.46806
## med eCDF diff..... 0.54766
## max eCDF diff..... 0.72924
##
## var ratio (Tr/Co)..... 0.13285
```



```

## T-test p-value..... < 2.22e-16
## Testing the balance for earnings in 1975
re75 <- balanceUV(treat$re75, no_treat$re75)
summary(re75)

## mean treatment..... 1532.1
## mean control..... 19063
## std mean diff..... -544.58
##
## mean raw eQQ diff..... 17978
## med raw eQQ diff..... 17903
## max raw eQQ diff..... 131511
##
## mean eCDF diff..... 0.46947
## med eCDF diff..... 0.53317
## max eCDF diff..... 0.77362
##
## var ratio (Tr/Co)..... 0.056057
## T-test p-value..... < 2.22e-16
## Testing the balance for unemployment status in 1974
u74 <- balanceUV(treat$u74, no_treat$u74)
summary(u74)

## mean treatment..... 0.70811
## mean control..... 0.086345
## std mean diff..... 136.39
##
## mean raw eQQ diff..... 0.62162
## med raw eQQ diff..... 1
## max raw eQQ diff..... 1
##
## mean eCDF diff..... 0.31088
## med eCDF diff..... 0.31088
## max eCDF diff..... 0.62176
##
## var ratio (Tr/Co)..... 2.6332
## T-test p-value..... < 2.22e-16
## Testing the balance for unemployment status in 1975
u75 <- balanceUV(treat$u75, no_treat$u75)
summary(u75)

## mean treatment..... 0.6
## mean control..... 0.1
## std mean diff..... 101.79
##
## mean raw eQQ diff..... 0.4973
## med raw eQQ diff..... 0
## max raw eQQ diff..... 1
##
## mean eCDF diff..... 0.25
## med eCDF diff..... 0.25
## max eCDF diff..... 0.5
##

```

```
## var ratio (Tr/Co)..... 2.6801
## T-test p-value..... < 2.22e-16
```

As we can see, there is a poor balance in the treatment.

- (e) Test whether the average post-treatment earnings difference between treated and control individuals is statistically different from zero. By doing this, are you assessing the average treatment effect? Why?

Testing:

```
test <- lm(re78 ~ treat, data = psid)
summary(test)

##
## Call:
## lm(formula = re78 ~ treat, data = psid)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21554  -9732   -866    7705   99620
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21553.9      303.6    70.98  <2e-16 ***
## treat        -15204.8     1154.6   -13.17  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15150 on 2673 degrees of freedom
## Multiple R-squared:  0.06092,    Adjusted R-squared:  0.06057
## F-statistic: 173.4 on 1 and 2673 DF,  p-value: < 2.2e-16
```

As we can see, the treatment is statistically significant in the regression, but we can not assess the real average treatment effect due to the poor balance.

- (f) How these results compare to the treatment effect obtained from the experimental data?

By doing a passable balance, the experimental data shows a positive impact on the average wage. On the other hand, former results show a negative one. Besides the selection bias in the treatment, all the explanations for why the treatment seems to be prejudicial are just conjectures.