Applied Microeconometrics - Assignment 3

Walter Verwer (589962) & Bas Machielsen (590049)

September 15, 2021

Construct a variable full-time equivalent for both waves, which is the number of full-time employees plus the number of part-time employees divided by two and also add the number of managers. I will simply refer to employees for this outcome variable.

```
dataset <- dataset %>%
    mutate(employees = (EMPFT + EMPPT)/2 + NMGRS,
           employees2 = (EMPFT2 + EMPPT2)/2 + NMGRS2,
           changeemployees = employees - employees2)
```

(i) Compute separately for New Jersey and Pennsylvania the average number of employees in both waves, and compute the difference-in-difference estimate

```
did <- dataset %>%
    group_by(STATE) %>%
    summarize(mean_before = mean(employees, na.rm = TRUE),
              mean_after = mean(employees2, na.rm = TRUE)) %>%
    mutate(STATE = if_else(STATE == 0, "PA", "NJ"))

did %>%
  knitr::kable(booktabs=T) %>%
  kableExtra::kable_styling(font_size = 7, latex_options = "hold_position")
```

STATE	mean_before	mean_after
PA	18.32468	17.38312
NJ	16.59439	16.80172

```
did$mean_after[2] - did$mean_after[1] - (did$mean_before[2] - did$mean_before[1])
```

```
## [1] 1.14889
```

Next repeat this, but only considering the restaurants that responded in both waves of the survey.

```
did2 <- dataset %>%
    group_by(STATE) %>%
    filter(!is.na(employees), !is.na(employees2)) %>%
    summarize(mean_before = mean(employees, na.rm = TRUE),
              mean_after = mean(employees2, na.rm = TRUE)) %>%
    mutate(STATE = if_else(STATE == 0, "PA", "NJ"))

did2$mean_after[2] - did2$mean_after[1] - (did2$mean_before[2] - did2$mean_before[1])
```

```
## [1] 1.169871
```

	NJ (N=309)		PA (N=75)	
	Mean	Std. Dev.	Mean	Std. Dev.
changeemployees	-0.110	5.104	1.060	6.087
NCALLS	1.214	1.464	0.747	0.960
WAGEST	4.609	0.343	4.630	0.358
INCTIME	17.905	10.625	19.279	13.183
FIRSTINC	0.228	0.110	0.210	0.096
BONUS	0.239	0.427	0.293	0.458
PCTAFF	49.157	34.789	45.571	36.935
MEALS	1.874	0.570	2.027	0.402
OPEN	8.100	2.182	7.807	2.164
HRSOPEN	14.398	2.818	14.513	2.960
PSODA	1.063	0.086	0.975	0.069
PFRY	0.941	0.103	0.843	0.089
PENTREE	1.360	0.657	1.232	0.635
NREGS	3.697	1.285	3.373	1.100
NREGS11	2.709	0.915	2.811	0.753

(ii) Estimate this model and next subsequently add characteristics of the restaurants observed in the first wave. But think carefully which characteristics can be included. How does the latter affect the estimate for the coefficient δ?

```
model1 <- lm(data = dataset,
   formula = changeemployees ~ STATE)

model2 <- update(model1, . ~ . + WAGE_ST)
model3 <- update(model2, . ~ . + INCTIME)
model4 <- update(model3, . ~ . + FIRSTINC)
model5 <- update(model4, . ~ . + BONUS)
model6 <- update(model5, . ~ . + PCTAFF)
model7 <- update(model6, . ~ . + MEALS)
model8 <- update(model7, . ~ . + OPEN)
model9 <- update(model8, . ~ . + HRSOPEN)
model10 <- update(model9, . ~ . + PSODA)
model11 <- update(model10, . ~ . + PFRY)
```

(iii) Provide a balancing table, i.e. show the sample mean of characteristics observed in the first survey separately for the restaurants in New Jersey and Pennsylvania. What is your opinion about the balancing table?

```
dataset %>%
    mutate(STATE = if_else(STATE == 0, "PA", "NJ")) %>%
    filter(!is.na(changeemployees)) %>%
    rename_with(.fn = ~ stringr::str_replace(.x, "_", "")) %>%
    select(changeemployees, STATE, NCALLS, WAGEST, INCTIME, FIRSTINC, BONUS, PCTAFF, MEALS,
           OPEN, HRSOPEN, PSODA, PFRY, PENTREE, NREGS, NREGS11) %>%
    modelsummary::datasummary_balance(formula = ~ STATE,
                                      dinm= TRUE,
                                      output = "latex",
                                      fmt = "%.3f",
                                      dinm_statistic = "p.value")
```

(iv) Check for the different characteristics if there is a common support for restaurants in New Jersey and Pennsylvania. And estimate a propensity score for being a restaurant in New Jersey.

	NJ				PA			
	Mean	SD	Boxplot	Histogram	Mean	SD	Boxplot	Histogram
NCALLS	1.21	1.46			0.75	0.96		
WAGEST	4.61	0.34			4.63	0.36		
INCTIME	17.91	10.63			19.28	13.18		
FIRSTINC	0.23	0.11			0.21	0.10		
BONUS	0.24	0.43			0.29	0.46		
PCTAFF	49.16	34.79			45.57	36.93		
MEALS	1.87	0.57			2.03	0.40		
OPEN	8.10	2.18			7.81	2.16		
HRSOPEN	14.40	2.82			14.51	2.96		
PSODA	1.06	0.09			0.98	0.07		
PFRY	0.94	0.10			0.84	0.09		
PENTREE	1.36	0.66			1.23	0.64		
NREGS	3.70	1.28			3.37	1.10		
NREGS11	2.71	0.92			2.81	0.75		

```
# Check for common support
dataset2 <- dataset %>%
    mutate(STATE = if_else(STATE == 0, "PA", "NJ")) %>%
    filter(!is.na(changeemployees)) %>%
    rename_with(.fn = ~ stringr::str_replace(.x, "_", "")) %>%
    select(changeemployees, STATE, NCALLS, WAGEST, INCTIME, FIRSTINC, BONUS, PCTAFF, MEALS,
           OPEN, HRSOPEN, PSODA, PFRY, PENTREE, NREGS, NREGS11)

emptycol = function(x) " "

boxplot1 <- lapply(dataset2 %>%
                     filter(STATE == "NJ") %>%
                  select(-STATE), na.omit) %>% lapply(scale)
boxplot2 <- lapply(dataset2 %>%
                     filter(STATE == "PA") %>%
                  select(-STATE), na.omit) %>% lapply(scale)

modelsummary::datasummary(
  data = dataset2,
   NCALLS + WAGEST + INCTIME + FIRSTINC + BONUS + PCTAFF + MEALS +
           OPEN + HRSOPEN + PSODA + PFRY + PENTREE + NREGS + NREGS11 ~
    STATE * (Mean + SD + Heading("Boxplot")*emptycol + Heading("Histogram")*emptycol)) %>%
  column_spec(column = 4, image = spec_boxplot(boxplot1)) %>%
  column_spec(column = 8, image = spec_boxplot(boxplot2)) %>%
  column_spec(column = 5, image = spec_hist(boxplot1)) %>%
  column_spec(column = 9, image = spec_hist(boxplot2))
```

```
# Estimate propensity score: p(X_i) = Pr(D_i=1|X_i), estimate via Logit.
logit <- glm(STATE ~ WAGE_ST + INCTIME + FIRSTINC + BONUS + PCTAFF + MEALS +
                  OPEN + HRSOPEN + PSODA + PFRY + PENTREE + NREGS + NREGS11 ,
                  family='binomial', data=dataset)
```

(v) Use propensity score matching to estimate the average treatment effect on the treated for the employment before and after the minimum wage increase in New Jersey, so on E_{0i} and E_{1i} separately.

(vi) Now use propensity score matching to estimate the average treatment effect on the treated on the change in employment in the restaurants, so $E_{1i} - E_{0i}$.

(vii) Now check the sensitivity of the propensity score matching estimate by also computing the weighting estimators for the average treatment effect on the treated.