Applied Microeconometrics - Assignment 3

Walter Verwer (589962) & Bas Machielsen (590049)

September 20, 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	23.33117	21.16558
NJ	20.43941	21.02743

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

```
## [1] 2.753606
```

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] 2.75
```

(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 <- lm(data = dataset,
             formula = changeemployees ~ STATE + SOUTHJ + CENTRALJ + SHORE + PA1)
model3 <- update(model1, . ~ . + NCALLS + WAGE_ST + INCTIME + FIRSTINC +
                   BONUS + MEALS + OPEN + HRSOPEN + PSODA + PFRY + NREGS + NREGS11)
model4 <- update(model2, . ~ . + NCALLS + WAGE_ST + INCTIME + FIRSTINC +
                   BONUS + MEALS + OPEN + HRSOPEN + PSODA + PFRY + NREGS + NREGS11)

models <- list(model1, model2, model3, model4)

stargazer(models, omit.stat = c("ll", "ser", "rsq"), df=F,
          omit = c("SOUTHJ", "CENTRALJ", "SHORE", "PA1"),
          add.lines = list(c("Region Dummies", "No", "Yes", "No", "Yes")),
          header = FALSE, font.size = "footnotesize", title='Estimated model for Q2.')
```

We want to isolate the effect of the minimum wage by attributing it to the coefficient belonging to STATE, which means we have to account for all possible sources of variation not due to the minimum wage. This also means we cannot control for PCTAFF, because this is the mechanism we care about: if we conditioned on this variable, that would absorb all variation due to the minimum wage policy changes and would change our interpretation of the STATE coefficient to a partial instead of a total effect and bias it towards zero.

(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.

```
# 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") %>%
```

Table 1: Estimated model for Q2.

	Dependent variable:			
	changeemployees			
	(1)	(2)	(3)	(4)
STATE	−2.750**	−1.785	−1.200	0.736
	(1.154)	(1.568)	(1.498)	(1.992)
NCALLS			−0.117	−0.136
			(0.371)	(0.371)
WAGE_ST			2.353	2.800*
			(1.512)	(1.543)
INCTIME			−0.072	−0.076
			(0.048)	(0.049)
FIRSTINC			−1.132	−1.460
			(5.133)	(5.164)
BONUS			0.209	0.230
			(1.203)	(1.214)
MEALS			−0.434	−0.269
			(0.998)	(1.001)
OPEN			−1.217*	−1.274*
			(0.724)	(0.747)
HRSOPEN			−0.576	−0.595
			(0.539)	(0.556)
PSODA			1.060	0.134
			(8.398)	(8.468)
PFRY			−6.126	−6.962
			(6.651)	(7.064)
NREGS			−0.544	−0.432
			(0.569)	(0.569)
NREGS11			0.458	0.419
			(0.788)	(0.790)
Constant	2.283**	0.970	16.376	14.339
	(1.036)	(1.399)	(16.219)	(17.705)
Region Dummies	No	Yes	No	Yes
Observations	384	384	303	303
Adjusted R^2	0.012	0.014	−0.002	0.005
F Statistic	5.675**	2.056*	0.962	1.091
Note:		*p<0.1; **p<0.05; ***p<0.01		

	NJ (N=309)		PA (N=75)	
	Mean	Std. Dev.	Mean	Std. Dev.
changeemployees	-0.467	8.452	2.283	10.854
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

```
                  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),
  title = 'Common support table.') %>%
  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))
```

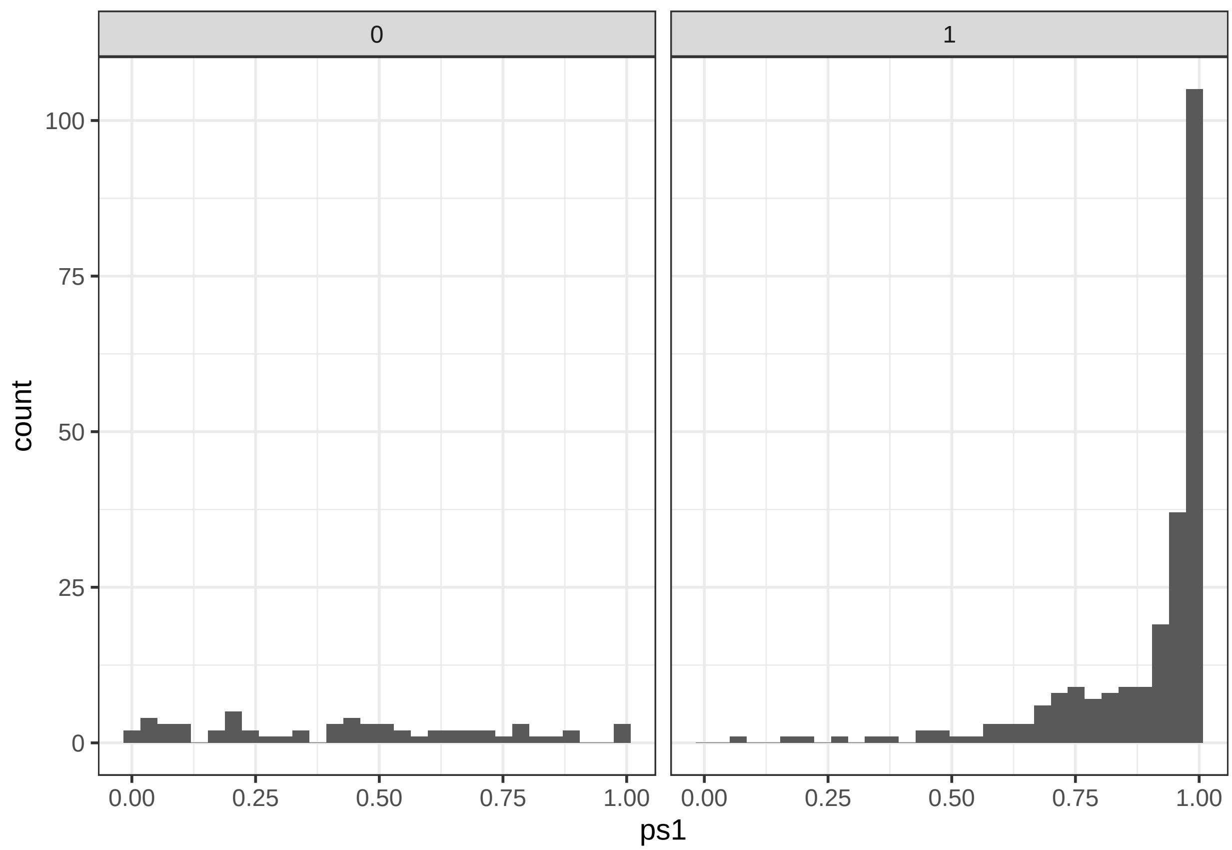
As indicated in the table, there is no common support for any of the variables, as we are dealing with continuous variables, so that the probability of realizing two zero outcomes is practically zero. We estimate two propensity scores, one extensive model, which sacrifices many observations, and one parsimonious model, which does not.

```
# Estimate propensity score
ps1 <- glm(STATE ~ WAGE_ST + INCTIME + FIRSTINC + BONUS + PCTAFF + MEALS + OPEN + HRSOPEN +
      PSODA + PFRY + PENTREE + NREGS + NREGS11 + NCALLS,
    data = dataset,
    family="binomial")

ps2 <- glm(STATE ~ MEALS + OPEN + HRSOPEN + PSODA + PFRY,
           data = dataset,
    family="binomial")

dataset <- dataset %>%
  modelr::add_predictions(ps1, type = "response") %>%
  rename("ps1" = "pred")

dataset <- dataset %>%
  modelr::add_predictions(ps2, type = "response") %>%
  rename("p2" = "pred")
```

```
dataset %>%
  ggplot(aes(x = ps1)) + geom_histogram()+ facet_wrap(~as.factor(STATE)) + theme_bw()
```

Table 2: Common support table.

	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		



(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.

We report the results of (v), (vi) and (viii) in table 3. We use the `MatchIt` package to estimate the propensity-score again and subsequently match using the nearest neighbor algorithm to compute E_{0i}:

```
matched_data1 <- MatchIt::matchit(STATE ~ MEALS + OPEN + HRSOPEN + PSODA + PFRY,
                  data = dataset %>%
                    select(employees, STATE, MEALS, OPEN, HRSOPEN, PSODA, PFRY) %>%
                    na.omit(),
                  method = "nearest") %>%
  match.data()

e_0i <- lm(employees ~ STATE + MEALS + OPEN + HRSOPEN + PSODA + PFRY, data = matched_data1)
```

And E_{1i}:

```
matched_data2 <- MatchIt::matchit(STATE ~ MEALS + OPEN + HRSOPEN + PSODA + PFRY,
                  data = dataset %>%
                    select(employees2, STATE, MEALS, OPEN, HRSOPEN, PSODA, PFRY) %>%
                    na.omit(),
                  method = "nearest") %>%
  match.data()

e_1i <- lm(employees2 ~ STATE + MEALS + OPEN + HRSOPEN + PSODA + PFRY, data = matched_data2)
```

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

```
matched_data3 <- MatchIt::matchit(STATE ~ MEALS + OPEN + HRSOPEN + PSODA + PFRY,
                  data = dataset %>%
                    select(changeemployees, STATE, MEALS, OPEN, HRSOPEN, PSODA, PFRY) %>%
                    na.omit(),
                  method = "nearest") %>%
  match.data()

ate <- lm(changeemployees ~ STATE + MEALS + OPEN + HRSOPEN + PSODA + PFRY, data = matched_data3)
```

```
stargazer(e_0i, e_1i, ate, header = F,
          omit.stat = c("ll", "ser", "rsq"), df = F,
          font.size = "footnotesize",
          label="tab:hoi",
          title='Application of propensity score matching to estimate the average treatment
          effect on the treated on the change in employments in the restaurants.')
```

(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.

We first calculate the nearest neighbour weighting estimate, using the `matched_data3` data.frame, which implements nearest-neighbor matching:

```
matched_data3 %>%
  arrange(subclass) %>%
  head(6) %>%
  kable(booktabs = TRUE, caption = "Nearest neigbour
                            weighting estimate.") %>%
  kableExtra::kable_styling(latex_options = c("hold_position", "scale_down"))
```

Table 3: Application of propensity score matching to estimate the average treatment effect on the treated on the change in employments in the restaurants.

	Dependent variable:		
	employees	employees2	changeemployees
	(1)	(2)	(3)
STATE	−1.262	1.410	−2.133
	(3.144)	(2.403)	(3.519)
MEALS	−0.095	1.360	−1.243
	(1.609)	(1.213)	(1.753)
OPEN	0.914	1.223	−0.436
	(0.981)	(0.765)	(1.103)
HRSOPEN	2.804***	2.518***	0.232
	(0.667)	(0.517)	(0.749)
PSODA	−18.029	−16.884*	−1.719
	(12.814)	(9.872)	(14.300)
PFRY	1.868	3.983	−2.654
	(8.838)	(6.914)	(9.983)
Constant	−7.829	−14.547	8.992
	(24.429)	(18.903)	(27.385)
Observations	150	150	146
Adjusted R^2	0.331	0.321	0.009
F Statistic	13.276***	12.756***	1.231
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Nearest neigbour weighting estimate.

changeemployees	STATE	MEALS	OPEN	HRSOPEN	PSODA	PFRY	distance	weights	subclass
-4.00	0	2	11	10.0	0.94	0.73	0.5233362	1	1
-0.50	1	1	9	12.5	1.12	1.02	0.9952999	1	1
3.50	0	3	7	16.0	1.00	0.87	0.4749829	1	2
-3.50	1	2	10	14.0	1.06	0.98	0.9940832	1	2
-16.00	0	2	6	16.0	0.94	0.90	0.2123310	1	3
22.75	1	1	7	15.0	1.17	0.95	0.9898831	1	3

```
treated_subjects <- sum(matched_data3$STATE)

differences <- matched_data3 %>%
  group_split(subclass) %>%
  map_dbl(.f =~ .x$changeemployees[2]- .x$changeemployees[1])

paste('The mean difference in employees is', mean(differences))
```

```
## [1] "The mean difference in employees is -3.1472602739726"
```

```
paste('The standard deviation is', sd(differences))
```

```
## [1] "The standard deviation is 14.6013481360335"
```

Now, we calculate the neighborhood matching weighting estimate, with $k = 4$, implemented again in the `MatchIt` package:

```
matched_data4 <- MatchIt::matchit(STATE ~ MEALS + OPEN + HRSOPEN + PSODA + PFRY,
                 data = dataset %>%
                   select(changeemployees, STATE, MEALS, OPEN, HRSOPEN, PSODA, PFRY) %>%
                   na.omit(),
                 method = "nearest",
                 replace = TRUE,
                 ratio = 4,
                 min.controls = 3) %>%
  get_matches()

matched_data4 %>%
  arrange(subclass) %>%
  head(6) %>%
  kable(booktabs=T, caption = "Neigborhood
                           weighting estimate, with $k=4$.") %>%
  kableExtra::kable_styling(latex_options = c("hold_position", "scale_down"))
```

Table 5: Neigborhood weighting estimate, with $k = 4$.

id	subclass	weights	changeemployees	STATE	MEALS	OPEN	HRSOPEN	PSODA	PFRY	distance
74	1	1.00	-12.0	1	2	7.0	16.0	1.06	0.95	0.9155793
34	1	0.25	-3.0	0	2	10.0	12.0	1.01	0.89	0.9235344
22	1	0.25	0.0	0	2	7.0	15.0	1.05	1.01	0.9002948
1	1	0.25	16.5	0	2	6.5	16.5	1.03	1.03	0.8989450
64	1	0.25	18.5	0	2	8.5	13.0	1.05	0.91	0.8924194
75	2	1.00	-6.5	1	2	7.0	14.5	1.06	0.95	0.8359729

```r
differences <- matched_data4 %>%
  group_split(subclass) %>%
  map_dbl(.f = ~ .x$changeemployees[1] - 0.25*(.x$changeemployees[2] +
                                                .x$changeemployees[3] +
                                                .x$changeemployees[4] +
                                                .x$changeemployees[5])
  )

paste('The mean difference in employees is', mean(differences))
```

```
## [1] "The mean difference in employees is -2.82320819112628"
```

```r
paste('The standard deviation is', sd(differences))
```

```
## [1] "The standard deviation is 10.0384946681803"
```

Finally, we manually implement the normal kernel density with $\Sigma = I$:

```r
library(mvtnorm)
```

```
## Warning: package 'mvtnorm' was built under R version 4.0.5
```

```r
normalkernel <- function(dataset){

  treated_outcomes <- dataset %>%
    filter(STATE ==1) %>%
    select(changeemployees)

  untreated_outcomes <- dataset %>%
    filter(STATE == 0) %>%
    select(changeemployees)

  treated_obs <- dataset %>%
    filter(STATE == 1) %>%
    select(MEALS, OPEN, HRSOPEN, PSODA, PFRY)

  untreated_obs <- dataset %>%
    filter(STATE == 0) %>%
    select(MEALS, OPEN, HRSOPEN, PSODA, PFRY)

  outcomes <- vector(length = nrow(treated_obs))
  w <- matrix(nrow = nrow(treated_obs), ncol = nrow(untreated_obs))

  for (i in 1:nrow(treated_obs)){

    for(j in 1:nrow(untreated_obs)){
      # Create the weight matrix
      w[i,j] <- mvtnorm::dmvnorm(as.numeric(untreated_obs[j,]),
                                 mean = as.numeric(treated_obs[i,]))
    }

    # Normalize the weights
  weightstotal <- sum(w[i,], na.rm = TRUE)
  w[i,] = w[i,] / weightstotal

  # Compute the estimated outcomes
  outcomes[i] <- treated_outcomes[i,] - sum(w[i,]*untreated_outcomes[,1], na.rm = TRUE)

  }

  outcomes <- unlist(outcomes)
```

```r
  print(paste('The estimated ATT is equal to:', mean(outcomes, na.rm = TRUE)))
  print(paste('The std. deviation is equal to:', sd(outcomes, na.rm = TRUE)))
  # Compute the average - final number
}
```

```r
normalkernel(dataset)
```

```
## [1] "The estimated ATT is equal to: -2.4880507067192"
## [1] "The std. deviation is equal to: 8.98329827881984"
```