# Applied Microeconometrics - Assignment 3

## Walter Verwer (589962) & Bas Machielsen (590049)

### September 17, 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 $\delta$?

```
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")
```

Table 1:

|  | *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. | Diff. in Means | p |
| changeemployees | -0.467 | 8.452 | 2.283 | 10.854 | 2.750 | 0.043 |
| NCALLS | 1.214 | 1.464 | 0.747 | 0.960 | −0.467 | 0.001 |
| WAGEST | 4.609 | 0.343 | 4.630 | 0.358 | 0.021 | 0.647 |
| INCTIME | 17.905 | 10.625 | 19.279 | 13.183 | 1.373 | 0.420 |
| FIRSTINC | 0.228 | 0.110 | 0.210 | 0.096 | −0.018 | 0.199 |
| BONUS | 0.239 | 0.427 | 0.293 | 0.458 | 0.054 | 0.357 |
| PCTAFF | 49.157 | 34.789 | 45.571 | 36.935 | −3.586 | 0.465 |
| MEALS | 1.874 | 0.570 | 2.027 | 0.402 | 0.153 | 0.008 |
| OPEN | 8.100 | 2.182 | 7.807 | 2.164 | −0.294 | 0.295 |
| HRSOPEN | 14.398 | 2.818 | 14.513 | 2.960 | 0.115 | 0.761 |
| PSODA | 1.063 | 0.086 | 0.975 | 0.069 | −0.087 | 0.000 |
| PFRY | 0.941 | 0.103 | 0.843 | 0.089 | −0.099 | 0.000 |
| PENTREE | 1.360 | 0.657 | 1.232 | 0.635 | −0.127 | 0.131 |
| NREGS | 3.697 | 1.285 | 3.373 | 1.100 | −0.324 | 0.029 |
| NREGS11 | 2.709 | 0.915 | 2.811 | 0.753 | 0.102 | 0.321 |

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") %>%
               select(-STATE), na.omit) %>% lapply(scale)
```

| | 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 | | |

```
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))
```

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

```
# 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()
```
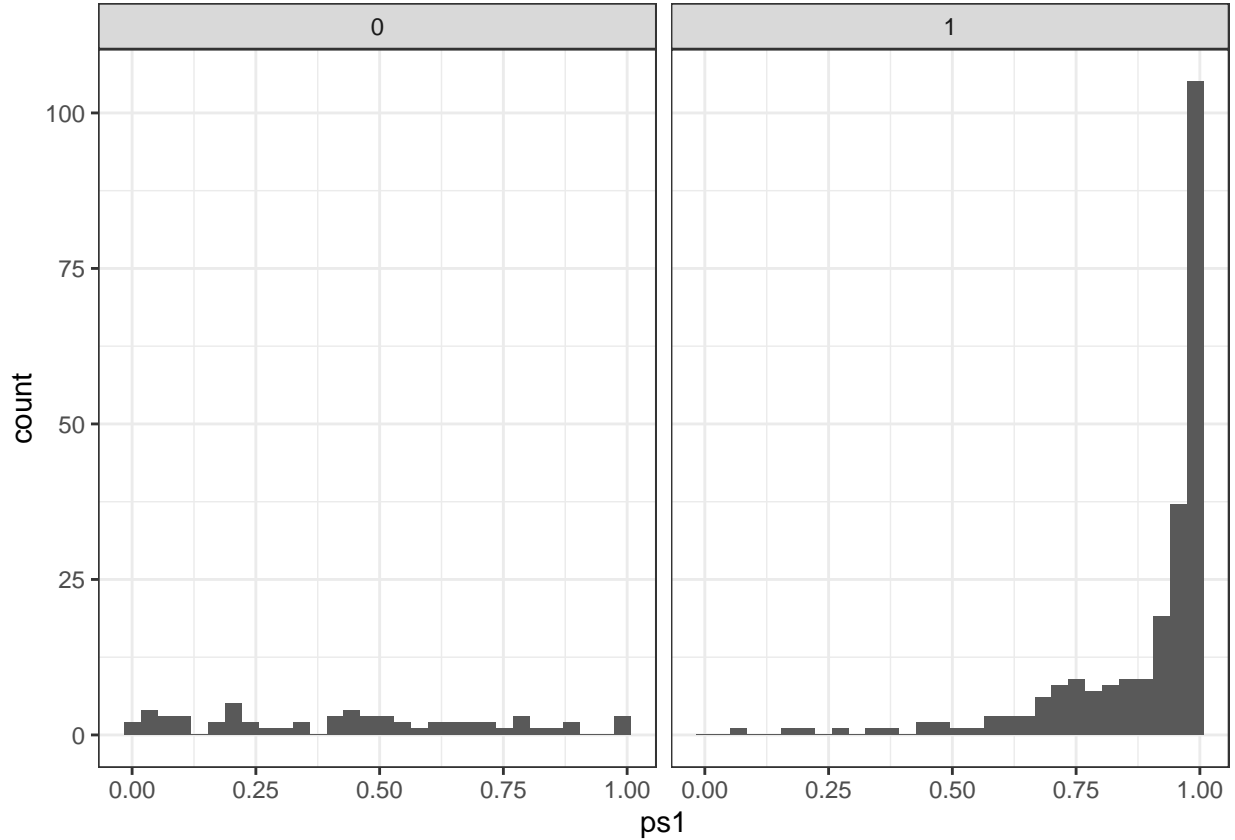
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 112 rows containing non-finite values (stat_bin).

(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 2. 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") %>%
```

5

```
  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")
```

Table 2:

| | *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:* | | | *$p<0.1$; **$p<0.05$; ***$p<0.01$ |

(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.