Assignment

483670 and 630516

September 18, 2022

“‘{ INFO, include=FALSE, results=‘asis’, echo=FALSE} Code Book for New Jersey-Pennsylvania Data Set

Note: there are 410 observations in the data set

```
    Column Location
```

Name: Start End Format Explanation ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff SHEET 1 3 3.0 sheet number (unique store id) CHAIN 5 5 1.0 chain 1=bk; 2=kfc; 3=roys; 4=wendys CO_OWNED 7 7 1.0 1 if company owned STATE 9 9 1.0 1 if NJ; 0 if Pa

Dummies for location: SOUTHJ 11 11 1.0 1 if in southern NJ CENTRALJ 13 13 1.0 1 if in central NJ NORTHJ 15 15 1.0 1 if in northern NJ PA1 17 17 1.0 1 if in PA, northeast suburbs of Phila PA2 19 19 1.0 1 if in PA, Easton etc SHORE 21 21 1.0 1 if on NJ shore

First Interview NCALLS 23 24 2.0 number of call-backs* EMPFT 26 30 5.2 # full-time employees EMPPT 32 36 5.2 # part-time employees NMGRS 38 42 5.2 # managers/ass’t managers WAGE_ST 44 48 5.2 starting wage ($/hr)INCTIME50545.1monthstousualfirstraiseFIRSTINC56605.2usualamountoffirstraise$(/hr) BONUS 62 62 1.0 1 if cash bounty for new workers PCTAFF 64 68 5.1 % employees affected by new minimum MEALS 70 70 1.0 free/reduced price code (See below) OPEN 72 76 5.2 hour of opening HRSOPEN 78 82 5.2 number hrs open per day PSODA 84 88 5.2 price of medium soda, including tax PFRY 90 94 5.2 price of small fries, including tax PENTREE 96 100 5.2 price of entree, including tax NREGS 102 103 2.0 number of cash registers in store NREGS11 105 106 2.0 number of registers open at 11:00 am

Second Interview TYPE2 108 108 1.0 type 2nd interview 1=phone; 2=personal STATUS2 110 110 1.0 status of second interview: see below DATE2 112 117 6.0 date of second interview MMDDYY format NCALLS2 119 120 2.0 number of call-backs* EMPFT2 122 126 5.2 # full-time employees EMPPT2 128 132 5.2 # part-time employees NMGRS2 134 138 5.2 # managers/ass’t managers WAGE_ST2 140 144 5.2 starting wage ($/hr)INCTIME21461505.1monthstousualfirstraiseFIRSTIN21521565.2usualamountoffirstraise$(/hr) SPECIAL2 158 158 1.0 1 if special program for new workers MEALS2 160 160 1.0 free/reduced price code (See below) OPEN2R 162 166 5.2 hour of opening HRSOPEN2 168 172 5.2 number hrs open per day PSODA2 174 178 5.2 price of medium soda, including tax PFRY2 180 184 5.2 price of small fries, including tax PENTREE2 186 190 5.2 price of entree, including tax NREGS2 192 193 2.0 number of cash registers in store NREGS112 195 196 2.0 number of registers open at 11:00 am

Codes:

Free/reduced Meal Variable: 0 = none 1 = free meals 2 = reduced price meals 3 = both free and reduced price meals

Second Interview Status 0 = refused second interview (count = 1) 1 = answered 2nd interview (count = 399) 2 = closed for renovations (count = 2) 3 = closed “permanently” (count = 6) 4 = closed for highway construction (count = 1) 5 = closed due to Mall fire (count = 1)

*Note: number of call-backs = 0 if contacted on first call

Q1 (Julian)

```r
NJemp1a <- mean(data$EMPFT[data$STATE == 1]) + 0.5 * mean(data$EMPPT[data$STATE ==
    1]) + mean(data$NMGRS[data$STATE == 1])
NJemp2a <- mean(data$EMPFT2[data$STATE == 1]) + 0.5 * mean(data$EMPPT2[data$STATE ==
    1]) + mean(data$NMGRS2[data$STATE == 1])
PAemp1a <- mean(data$EMPFT[data$STATE == 0]) + 0.5 * mean(data$EMPPT[data$STATE ==
    0]) + mean(data$NMGRS[data$STATE == 0])
PAemp2a <- mean(data$EMPFT2[data$STATE == 0]) + 0.5 * mean(data$EMPPT2[data$STATE ==
    0]) + mean(data$NMGRS2[data$STATE == 0])
EffNJa <- NJemp2a - NJemp1a
EffPAa <- PAemp2a - PAemp1a
DiffDiffa <- EffNJa - EffPAa
DiffDiffa
```

```
## [1] 2.75
```

```r
data1 <- subset(data, STATUS2 == 1)

NJemp1b <- mean(data1$EMPFT[data1$STATE == 1]) + 0.5 * mean(data1$EMPPT[data1$STATE ==
    1]) + mean(data1$NMGRS[data1$STATE == 1])
NJemp2b <- mean(data1$EMPFT2[data1$STATE == 1]) + 0.5 * mean(data1$EMPPT2[data1$STATE ==
    1]) + mean(data1$NMGRS2[data1$STATE == 1])

PAemp1b <- mean(data1$EMPFT[data1$STATE == 0]) + 0.5 * mean(data1$EMPPT[data1$STATE ==
    0]) + mean(data1$NMGRS[data1$STATE == 0])
PAemp2b <- mean(data1$EMPFT2[data1$STATE == 0]) + 0.5 * mean(data1$EMPPT2[data1$STATE ==
    0]) + mean(data1$NMGRS2[data1$STATE == 0])


EffNJb <- NJemp2b - NJemp1b
EffPAb <- PAemp2b - PAemp1b
DiffDiffb <- EffNJb - EffPAb
DiffDiffb
```

```
## [1] 2.721919
```

```r
data <- data1
```

The mean number of employees in NJ are 20.43 and 20.9 for the first and second wave respectively. In Pa the employee numbers were, respectively 23.38 and 21.1 for the first and second wave. The naive treatment effect is then 0.47 for NJ and -2.28 for Pa. Taking the diff-in-diff estimator of this we get 2.75. If we adjust the data set to only include those who participated both times this effect changes to 3, broken down into an effect of 0.66 for NJ and -2.06 for Pa. The treatment effect is thus very similar, but decreases slightly in magnitude.

Q2

We first estimate the model with a dummy variable indicating whether the restaurant is in New Jersey (D = 1) or in Pennsylvenia (D = 0). We get an treatment effect of 2.722 which means that thanks to the minimum wage increase around 2.7 workers more will be employed in a restaurant in New Jersey. The effect is significant at 5%.

Next, we want to include controls. However, in our view it is not sensible the include solely characteristics from the first wave. Instead, it would make more sense to include the difference of the characterstics as a control. To explain our point, suppose the model

$$E_{igt} = \gamma_g + \lambda_t + \delta NJ_{igt} + \epsilon_{igt},$$

where E_{igt} is the amount of workers employed at restaurant i, in state g, and time t. γ_g are state specific effects, λ_t are time effects, NJ_i is a dummy variable indicating whether the restaurant is in New Jersey. We have two states, $g = \{NJ, PA\}$ and two time periods, $t = \{0, 1\}$, so we can rewrite the model

$$E_{i,NJ,0} = \gamma_{NJ}+\lambda_0+\epsilon_{i,NJ,0}E_{i,NJ,1} = \gamma_{NJ}+\lambda_1+\delta+\epsilon_{i,NJ,1}E_{i,PA,0} = \gamma_{PA}+\lambda_0+\epsilon_{i,PA,0}E_{i,PA,1} = \gamma_{PA}+\lambda_1+\epsilon_{i,PA,1}$$

Next, we take the differences

$$E_{i,NJ,1} - E_{i,NJ,0} = \lambda_1 - \lambda 0 + \delta + \epsilon_{i,NJ,1} - \epsilon_{i,NJ,0}E_{i,PA,1} - E_{i,PA,0} = \lambda_1 - \lambda 0 + \epsilon_{i,PA,1} - \epsilon_{i,PA,0}$$

where we define $\alpha = \lambda_1 - \lambda_0$. These two equations we can rewrite into our model $E_{1i} - E_{0i} = \alpha + \delta NJ_i + U_i$. Now suppose we add level controls, X_{igt} to our model

$$E_{igt} = \gamma_g + \lambda_t + \delta NJ_{igt} + \beta X_{igt} + \epsilon_{igt},$$

Again, we rewrite the model. At this point the question arises whether we add controls for both time periods or only the first one. While it makes sense to add controls to reduce biases, we do not think it would make sense to add the controls only at $t = 0$.

$$E_{i,NJ,0} = \gamma_{NJ}+\lambda_0+\beta X_{i,NJ,0}+\epsilon_{i,NJ,0}E_{i,NJ,1} = \gamma_{NJ}+\lambda_1+\delta+\beta X_{i,NJ,1}+\epsilon_{i,NJ,1}E_{i,PA,0} = \gamma_{PA}+\lambda_0+\beta X_{i,PA,0}+\epsilon_{i,PA,0}E_{i,PA,1} =$$

However, if we use this model setup, we would not include level variables but the difference of the variable for the control. This makes also conceptually sense. Suppose a free meal policy has a significant effect on the amount of employed workers, and we want to know whether there are other variables that affect the difference in employment between $t = 0$, and $t = 1$. Now, we have several restaurants that change their meal policies, which affects the employment rate. Assuming such a situation it would make no sense to include only the variable meal at $t = 0$ but the difference between $t = 1$ and $t = 0$ as otherwise we do not measure the effect of the change in the meal policy on the change of the amount of workers.\

As an example, we include the difference in the hours of opening. We can see, that the state effect slightly decreases but the significance stays at 5%. Moreover, the difference in hours of opening has a negative effect on the difference in employment, which means that when a restaurant opens up 1 hour later than in the previous period, then around 1.7 workers less will be employed at the restaurant. The estimation is significant at 1%. A variable that we can include and that has not to be differenced is the amount of employees affected by the minimum (as it is zero in the second period). It is sensible to include this control as the amount of affected employees potentially can impact the decision of managers on how many persons to employ after the policy is introduced. When we include only the amount of affected employees as a control, the state effect decreases to 2.466 but is still significant at 5%. Furthermore, the number of affected employees has a slightly positive effect with 0.034 and is significant at 5%. The adjusted R^2 is hight than in the basismodel, and the residual standard error slightly lower. One can also combine both controls into one model. In this case the state effect decreases to 2.452 and is still siginificant at 5%. The controls are significant and of similar dimension as in the previous models. The adjsuted R^2 is higher than in the previous models, and the residual standard error lower. However, as the question asked to estimate a model using controls from the first survey, we will stick to the model including the number of affected employees as a control variable. All other variables should be included as a difference when being used as a control variable.

```r
data$EMP1 <- data$EMPFT + 0.5 * data$EMPPT + data$NMGRS
data$EMP2 <- data$EMPFT2 + 0.5 * data$EMPPT2 + data$NMGRS2
data$difE <- data$EMP2 - data$EMP1
data$difOpen <- data$OPEN2R - data$OPEN
m1 <- lm(difE ~ STATE, data)
m2 <- lm(difE ~ STATE + difOpen, data)
m3 <- lm(difE ~ STATE + PCTAFF, data)
m4 <- lm(difE ~ STATE + difOpen + PCTAFF, data)
dif_eff1 <- m1$coefficients[2]
stargazer(m1, m2, m3, m4, column.labels = c("", "Controls"), type = "text",
    title = "Minimum Wage", header = FALSE, label = "tab:reg1")
```

```
##
## Minimum Wage
## ==========================================================================================================
##                                                         Dependent variable:
##                        -----------------------------------------------------------------------------------
##                                                                 difE
##                                                  Controls
##                                   (1)                (2)                   (3)                   (4)
## ----------------------------------------------------------------------------------------------------------
## STATE                           2.722**            2.689**               2.466**               2.452*
##                                 (1.153)            (1.146)               (1.184)               (1.175
##
## difOpen                                           -1.682***                                   -1.685*
##                                                    (0.631)                                     (0.63
##
## PCTAFF                                                                   0.034**               0.033*
##                                                                          (0.014)               (0.01
##
## Constant                       -2.064**            -1.939*              -3.597***             -3.425*
##                                 (1.034)            (1.028)               (1.219)               (1.21
##
## ----------------------------------------------------------------------------------------------------------
## Observations                      378                377                   339                   338
## R2                               0.015              0.033                 0.032                 0.05
## Adjusted R2                      0.012              0.028                 0.026                 0.04
## Residual Std. Error        8.897 (df = 376)     8.837 (df = 374)      8.768 (df = 336)      8.701 (df
## F Statistic              5.570** (df = 1; 376) 6.357*** (df = 2; 374) 5.547*** (df = 2; 336) 6.115*** (df
## ==========================================================================================================
## Note:                                                                              *p<0.1; **p<0.05;
```

Q3

Media_control1 is the median for Pennsylvenia and Media_trat1 is the median for New Jersey. We ignore the location dummies as they naturally results in significant differences among the states. For the other variables, we can see a significant difference in full-time employees (EMPFT) in the first survey year as restaurants in New Jersey seem to have more employees than restaurants in Pennsylvenie according to the median statistics. However, an insightful aspect is that in the second survey the median in Pennsylvenia is considerable higher (EMPFT = 1.02, EMPFT2 = 7.78) which is why the difference between the states with respect to the number of full-time employees is not significant anymore in the second survey. Another significant difference between the states can be observed for the usual amount of first raise ($/hr) in the second survey which is higher in New Jersey than in Pennsylvenia. Comparing it to the first survey, we can see that before the usual amount of first raise ($/hr) was higher in both states but it decreased more for Pennsylvenia. Generally, participating restaurants in Pennsylvenia have substantial higher prices across all products in both survey. Moreover, the starting wage in the second survey is significantly higher for New Jersey. However, this does not only results from an increase of the Median for restaurants in New Jersey but also from a decrease of around 0.03 for restaurants in Pennsylvenia.\

All in all, there are considerable differences between the states. However, not every significant difference can be observed in both surveys. Especially, when it comes to the starting wage and the amount of first raise, where significant differences between the states can be only observed for the second survey, we must consider that they result from the state effects. Nonetheless, one aspects that worries us is the significant difference in full-time employees that is lost in the second survey. Moreover, it seems that there are considerable price differences across the states but it is not clear whether it is due to general price differences between the states or due to more expensive restaurants answering the survey in Pennsylvenia.

```
data3 <- subset(data, select = -c(STATUS2))
balance <- balance_table(data3, "STATE")
knitr::kable(balance, caption = "Balance Table")
```

Q4 (Julian)

To estimate the propensity score we run a probit model estimating the probability of being in a given state. Table **??** calculates the probability of each restaurant being in NJ as opposed to Pa.

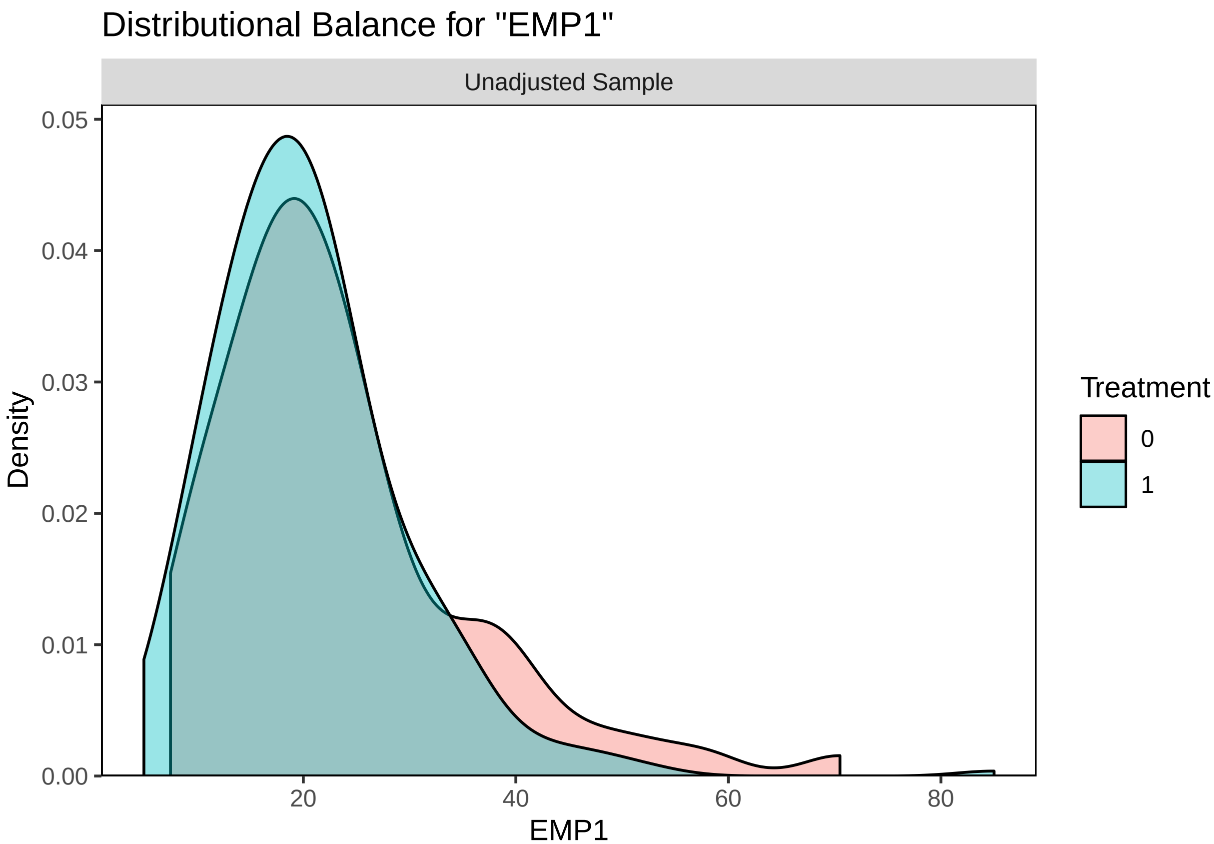
We do not include location dummies in the estimation as these affect the treatment but not the outcome variable, hence they can be used as instruments for istance. We use all variables that jointly determine the treatment and the outcome.

```
dataNJ <- subset(data, STATE == 1)
dataPA <- subset(data, STATE == 0)
bal.plot(data, treat = data$STATE, var.name = "EMP1")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```

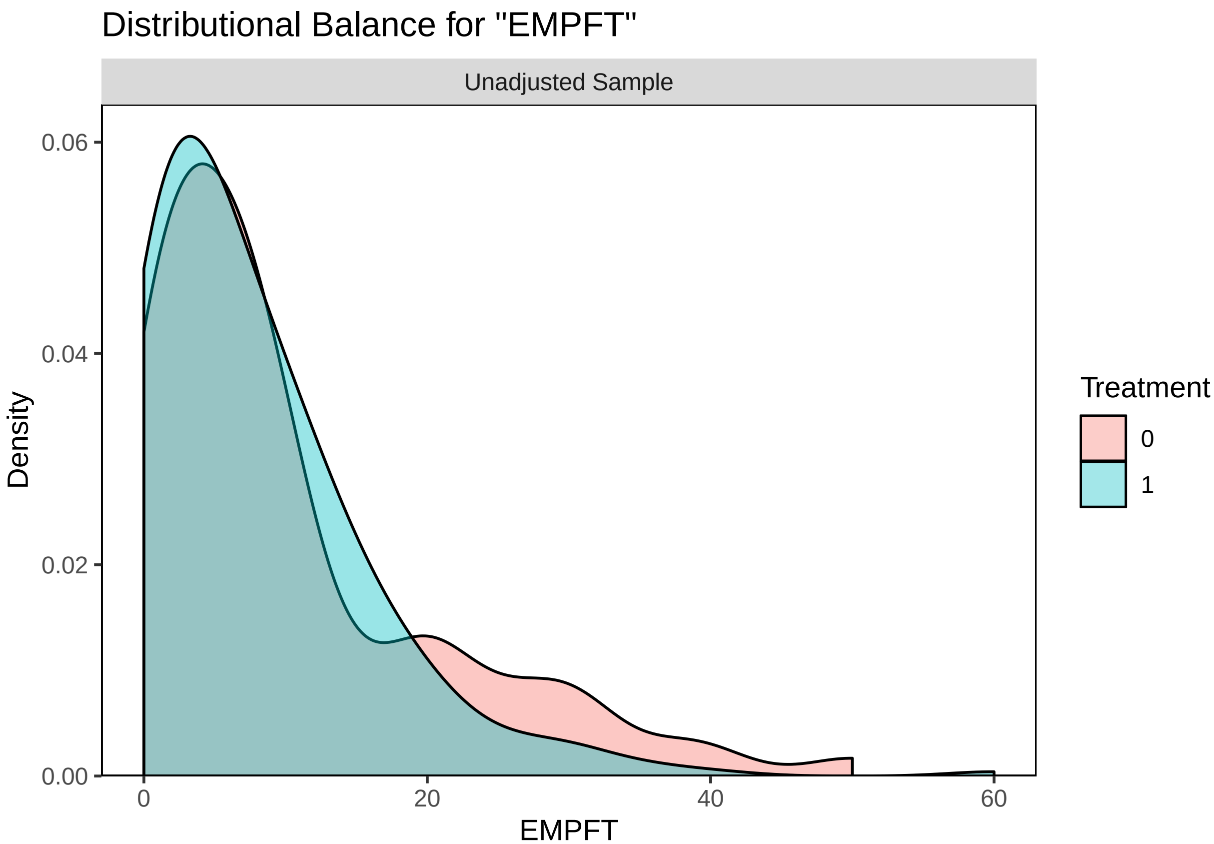
Table 1: Balance Table

variables1	Media_control1	Media_trat1	p_value1
BONUS	2.972973e-01	2.434211e-01	0.3624466
CENTRALJ	0.000000e+00	1.875000e-01	0.0000000
CHAIN	2.121622e+00	2.098684e+00	0.8782334
CO_OWNED	3.513514e-01	3.453947e-01	0.9238736
DATE2	1.121285e+05	1.120950e+05	0.9261648
difE	-2.064189e+00	6.577303e-01	0.0450104
difOpen	7.432430e-02	5.858090e-02	0.8348549
EMP1	2.344595e+01	2.058322e+01	0.0593918
EMP2	2.138176e+01	2.124095e+01	0.8956520
EMPFT	1.020270e+01	7.751645e+00	0.0708302
EMPFT2	7.777027e+00	8.496711e+00	0.5118295
EMPPT	1.937838e+01	1.884211e+01	0.6749107
EMPPT2	1.983108e+01	1.850493e+01	0.3034505
FIRSTIN2	1.844444e-01	2.246032e-01	0.0012576
FIRSTINC	2.079688e-01	2.297091e-01	0.1133117
HRSOPEN	1.453378e+01	1.439638e+01	0.7197910
HRSOPEN2	1.463514e+01	1.438696e+01	0.5044687
INCTIME	1.936957e+01	1.793214e+01	0.4053026
INCTIME2	2.115152e+01	2.235135e+01	0.4757115
MEALS	2.027027e+00	1.871711e+00	0.0076043
MEALS2	1.905405e+00	1.733553e+00	0.0035400
NCALLS	7.567568e-01	1.213816e+00	0.0013335
NCALLS2	1.677419e+00	2.281818e+00	0.0262003
NMGRS	3.554054e+00	3.410526e+00	0.3077894
NMGRS2	3.689189e+00	3.491776e+00	0.2040121
NORTHJ	0.000000e+00	5.263158e-01	0.0000000
NREGS	3.378378e+00	3.705686e+00	0.0296376
NREGS11	2.821918e+00	2.710884e+00	0.2833660
NREGS112	2.575343e+00	2.683849e+00	0.2735462
NREGS2	3.479452e+00	3.671186e+00	0.2301576
OPEN	7.797297e+00	8.110197e+00	0.2704106
OPEN2R	7.871622e+00	8.172442e+00	0.2879269
PA1	4.594595e-01	0.000000e+00	0.0000000
PA2	5.405405e-01	0.000000e+00	0.0000000
PCTAFF	4.478261e+01	4.866407e+01	0.4289599
PENTREE	1.236389e+00	1.366122e+00	0.1278442
PENTREE2	1.190986e+00	1.413265e+00	0.0061747
PFRY	8.418056e-01	9.414533e-01	0.0000000
PFRY2	8.598551e-01	9.594178e-01	0.0000000
PSODA	9.741667e-01	1.062349e+00	0.0000000
PSODA2	9.740541e-01	1.062925e+00	0.0000000
SHEET	3.759459e+02	2.163125e+02	0.0000000
SHORE	0.000000e+00	1.052632e-01	0.0000000
SOUTHJ	0.000000e+00	2.861842e-01	0.0000000
SPECIAL2	2.328767e-01	2.046980e-01	0.6097328
TYPE2	1.067568e+00	1.065789e+00	0.9566654
WAGE_ST	4.632254e+00	4.612222e+00	0.6726108
WAGE_ST2	4.617246e+00	5.081200e+00	0.0000000



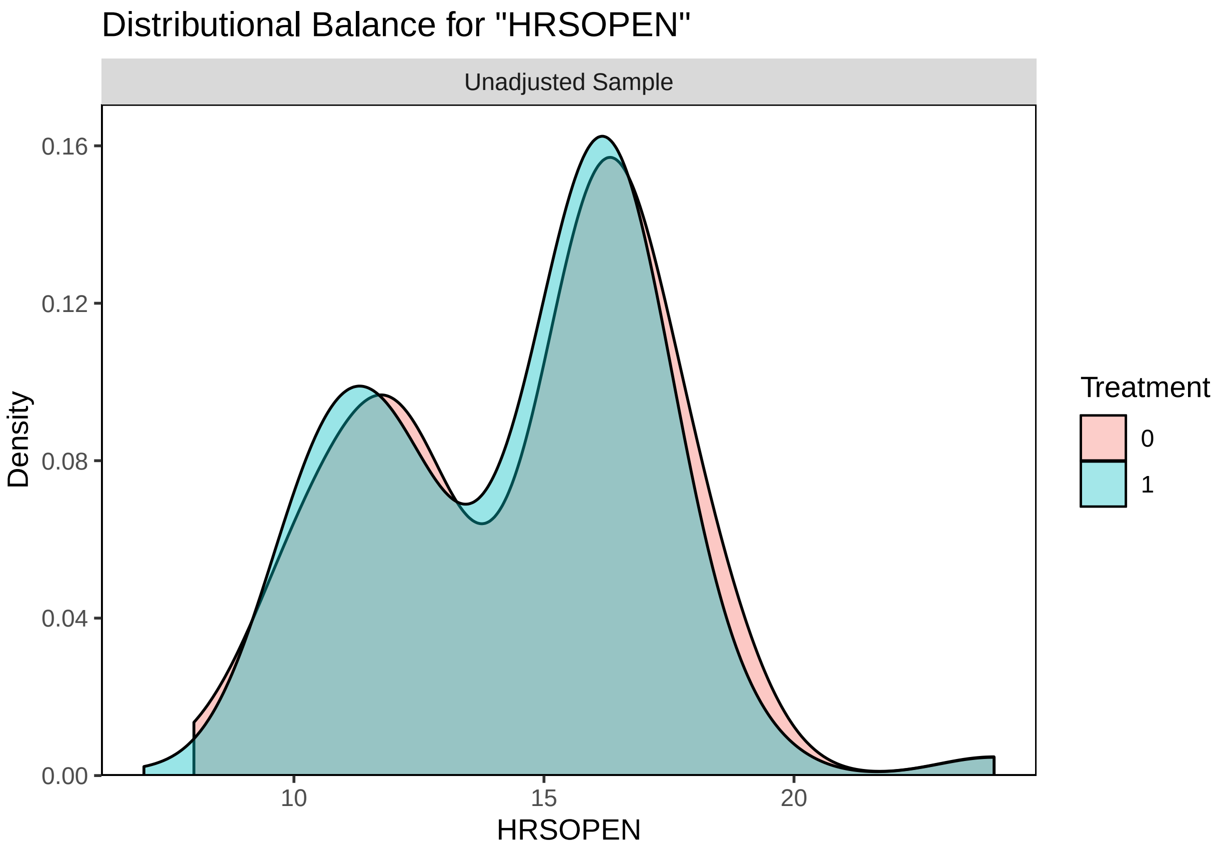

```
bal.plot(data, treat = data$STATE, var.name = "EMPFT")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```



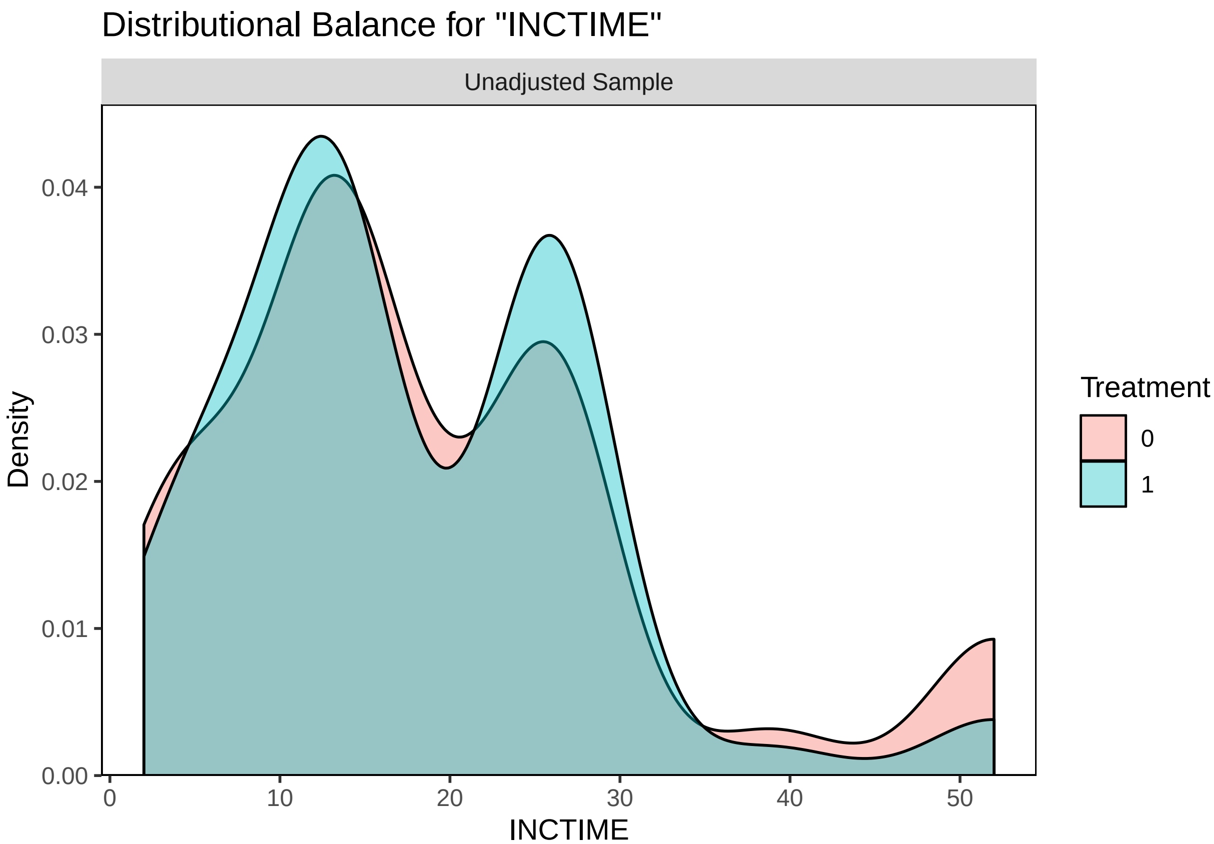

```
bal.plot(data, treat = data$STATE, var.name = "HRSOPEN")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```



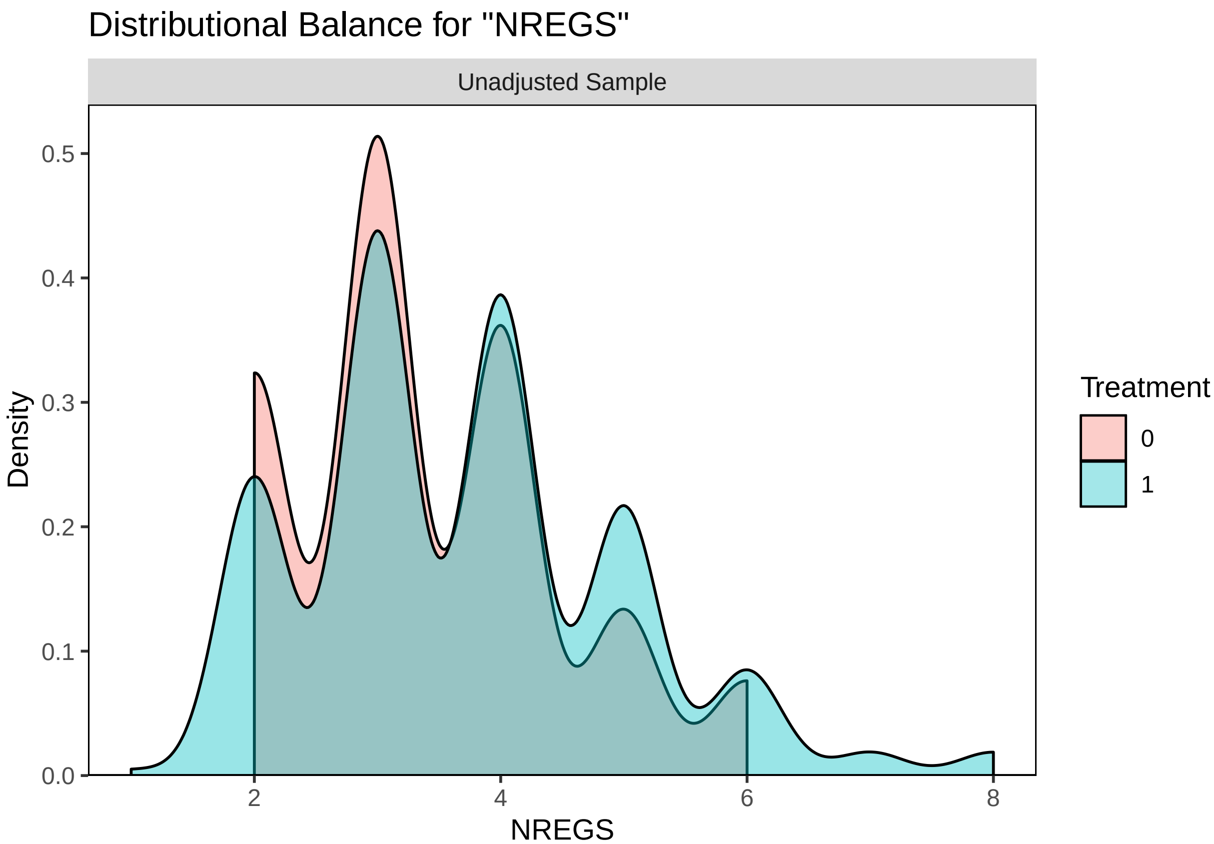

```r
bal.plot(data, treat = data$STATE, var.name = "INCTIME")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```



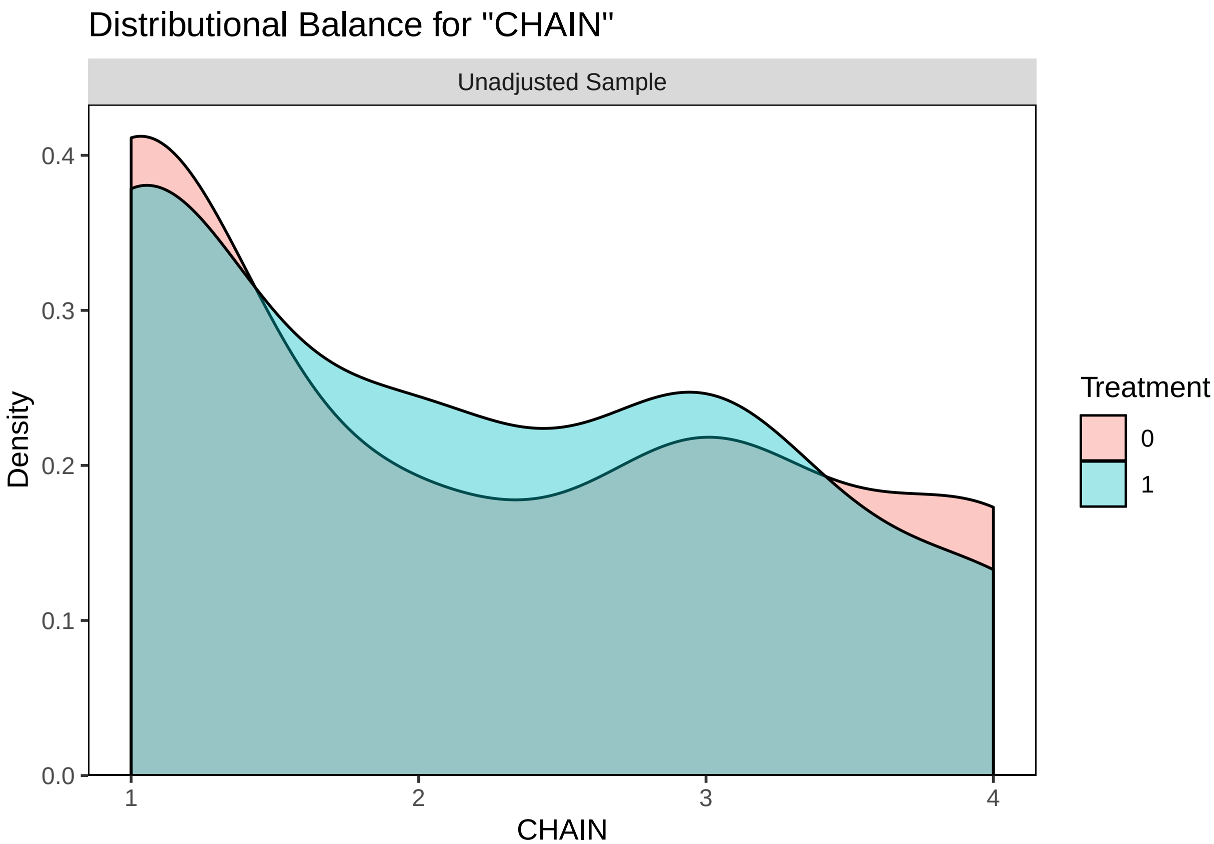

```
bal.plot(data, treat = data$STATE, var.name = "NREGS")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```



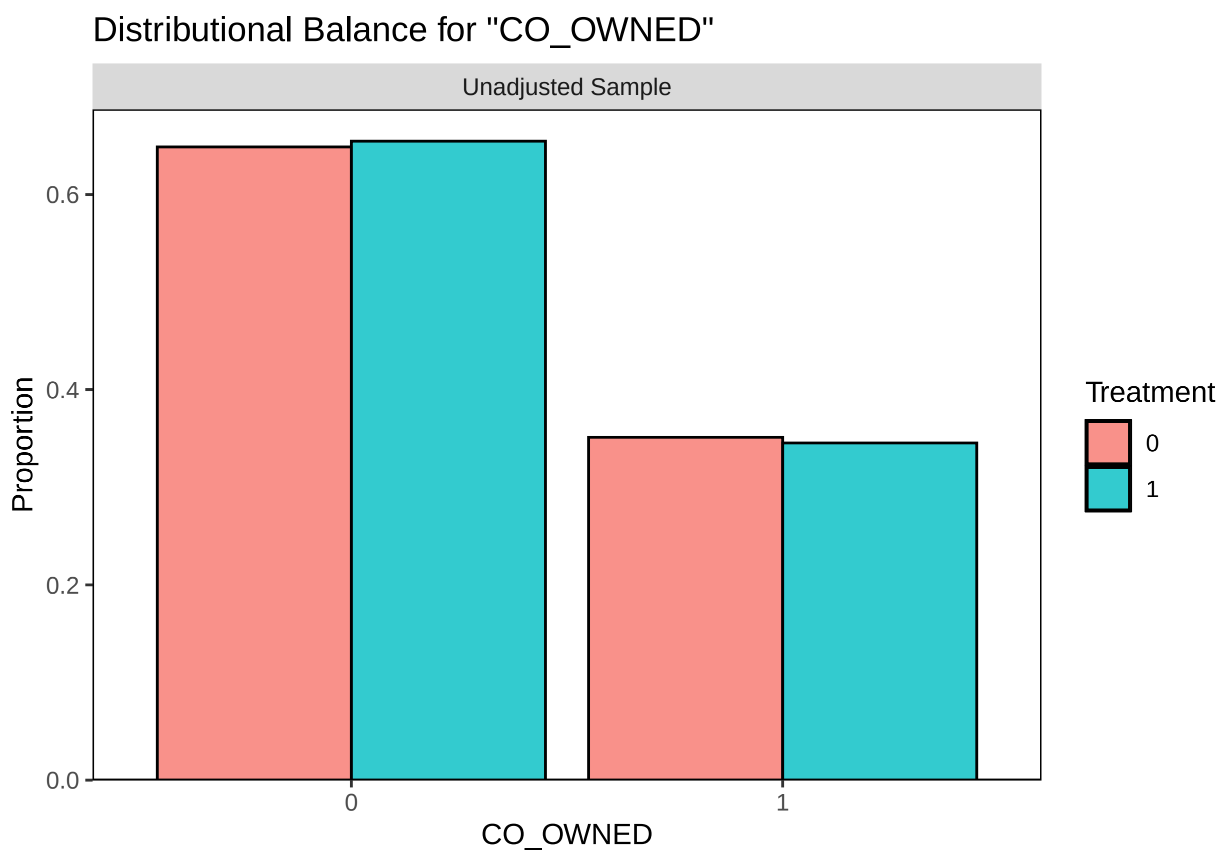

```
bal.plot(data, treat = data$STATE, var.name = "CHAIN")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```



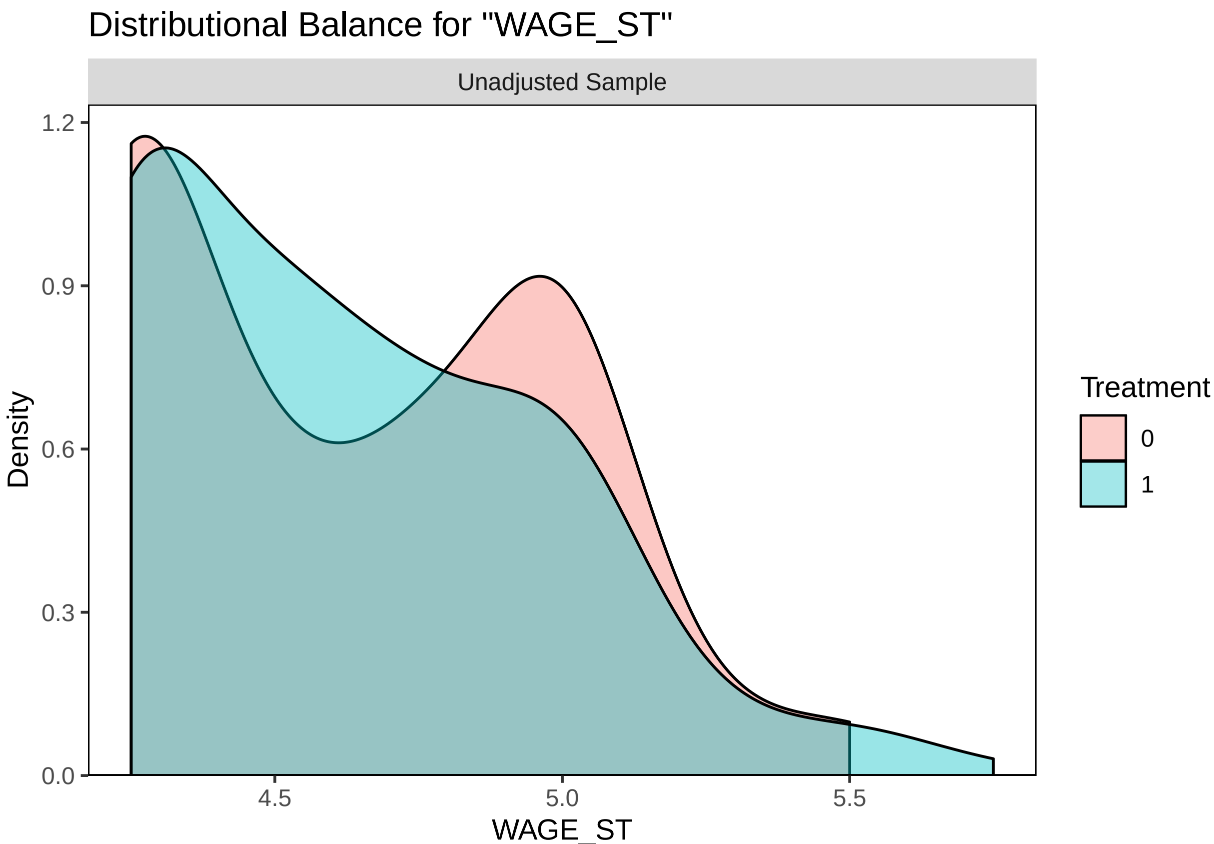

```
bal.plot(data, treat = data$STATE, var.name = "CO_OWNED")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```



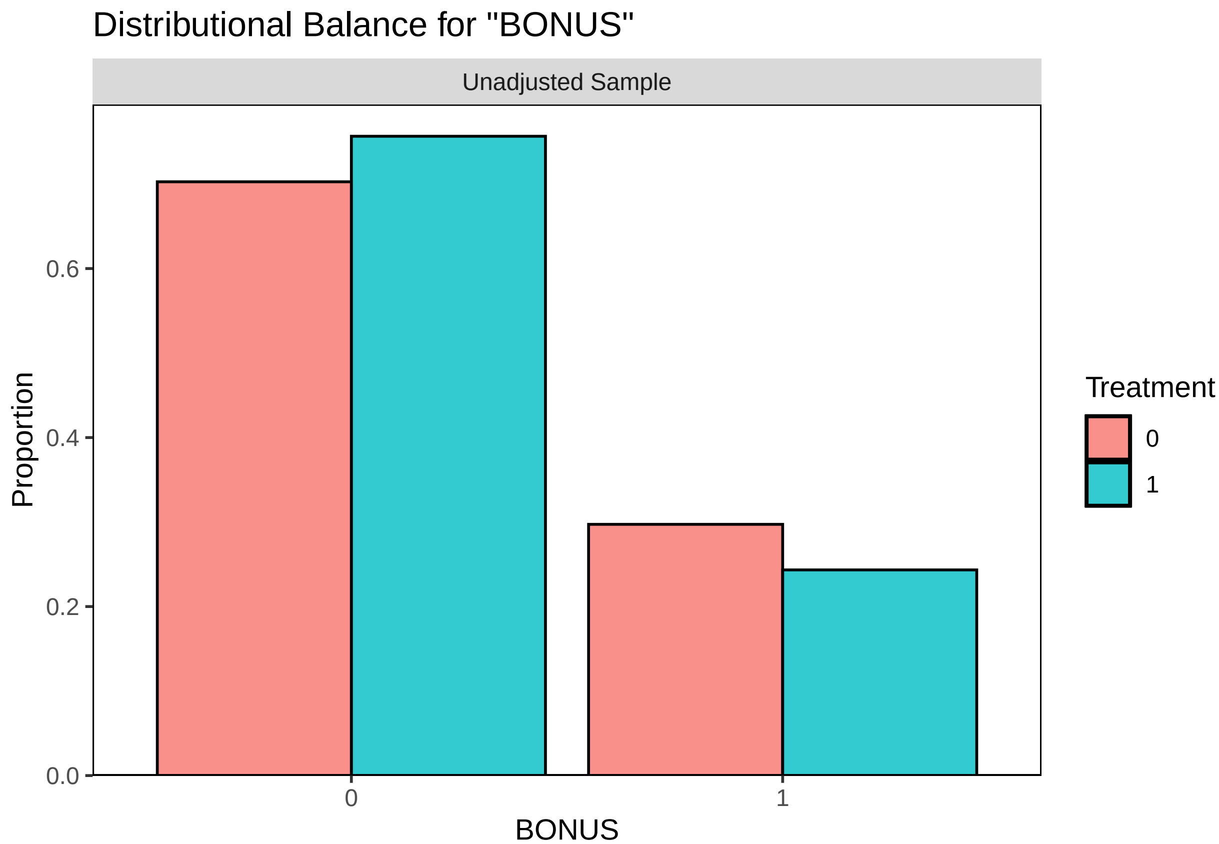

```
bal.plot(data, treat = data$STATE, var.name = "WAGE_ST")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```



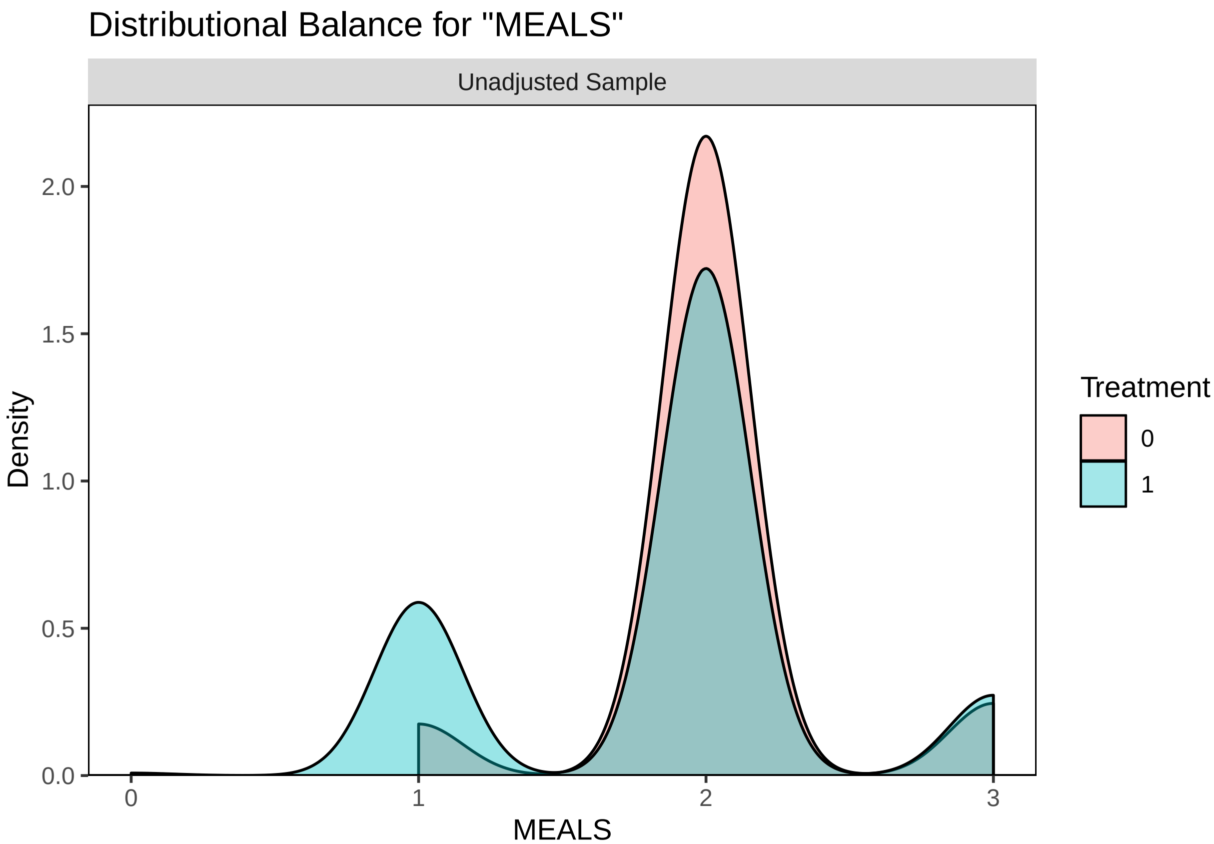

```
bal.plot(data, treat = data$STATE, var.name = "BONUS")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```



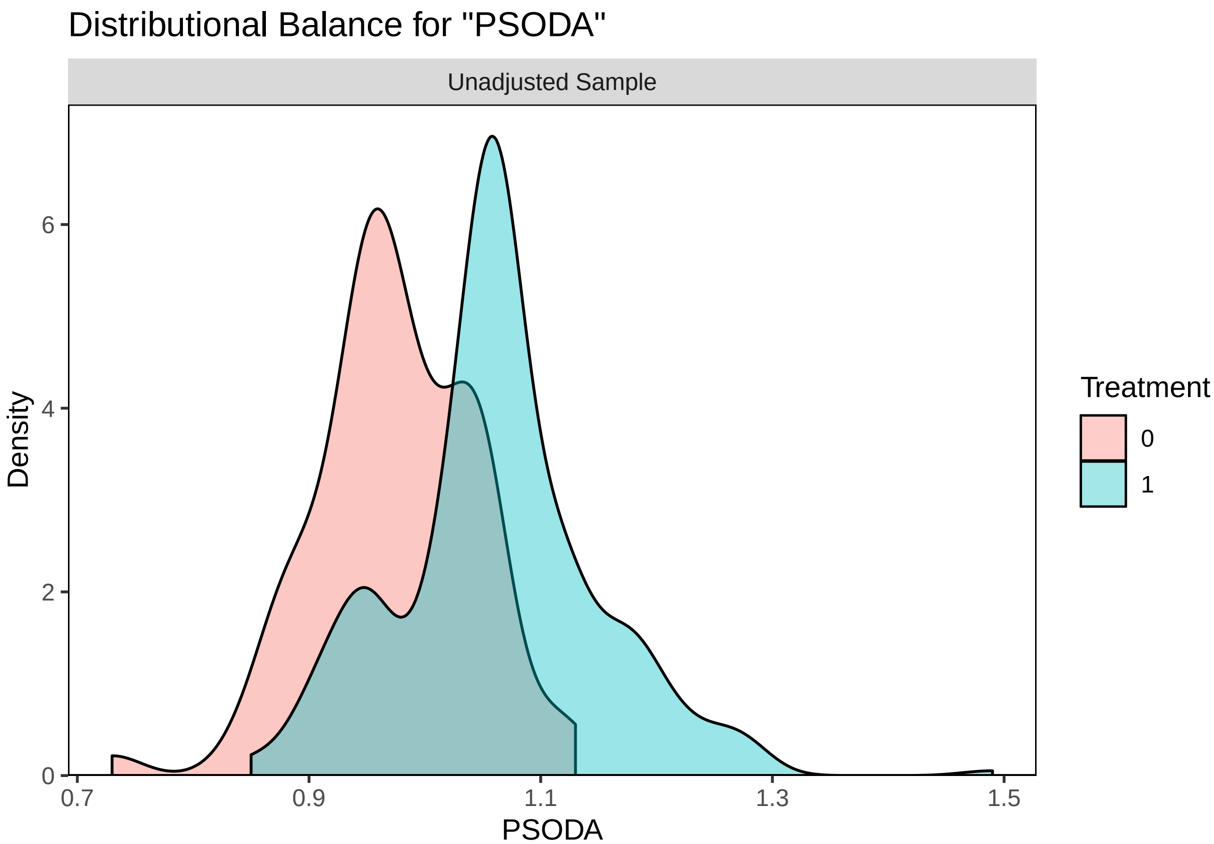

```
bal.plot(data, treat = data$STATE, var.name = "MEALS")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```




```r
bal.plot(data, treat = data$STATE, var.name = "PSODA")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```



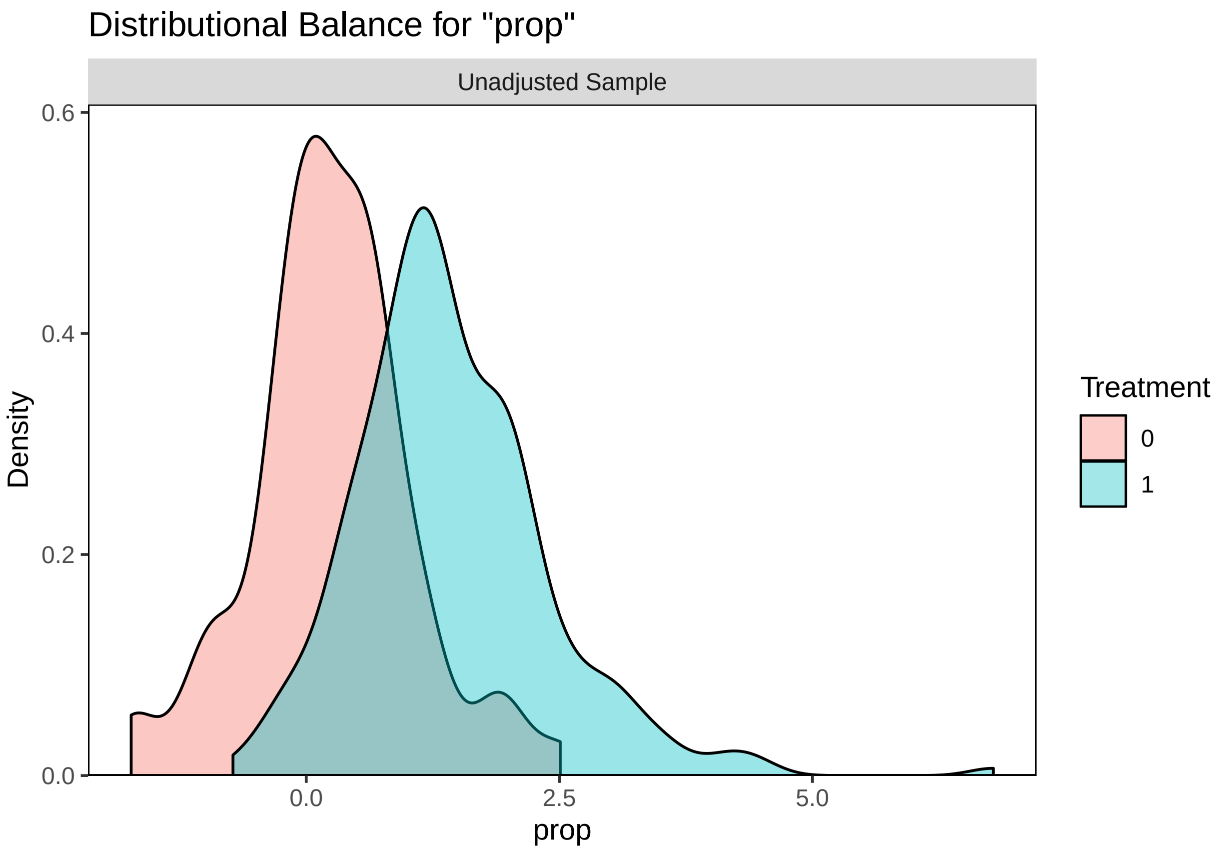

```
probit <- glm(STATE ~ HRSOPEN + INCTIME + NREGS + CHAIN + CO_OWNED +
    WAGE_ST + BONUS + MEALS + PSODA, family = binomial(link = "probit"),
    data = data)
data$prop <- predict(probit, type = "response", data)
stargazer(probit, column.labels = c("", ""), type = "text", title = "Minimum Wage",
    header = FALSE, label = "tab:probit")
```

```
##
## Minimum Wage
## =============================================
##                       Dependent variable:
##                     ---------------------------
##                                STATE
##
## ---------------------------------------------
## HRSOPEN                      -0.149***
##                               (0.045)
##
## INCTIME                       -0.005
##                               (0.008)
##
## NREGS                         -0.005
##                               (0.088)
##
## CHAIN                         -0.067
##                               (0.090)
```

```
##
## CO_OWNED                       -0.353
##                                (0.219)
##
## WAGE_ST                        -0.214
##                                (0.290)
##
## BONUS                          -0.019
##                                (0.221)
##
## MEALS                         -0.485***
##                                (0.187)
##
## PSODA                         11.584***
##                                (1.643)
##
## Constant                      -6.458***
##                                (1.758)
##
## ---------------------------------------------
## Observations                     325
## Log Likelihood                -117.994
## Akaike Inf. Crit.              255.987
## =============================================
## Note:             *p<0.1; **p<0.05; ***p<0.01
```

```
dataNJ <- subset(data, STATE == 1)
dataPA <- subset(data, STATE == 0)
bal.plot(data, treat = data$STATE, var.name = "prop")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```



Q5

Using the nearest neighbour method, we use the propensity score to match the restaurants and estimate the ATET. The graph shows that the there is an overlapping support between the propensity scores of the matched restaurants. We then estimate two models, one for the employment rate before and one after the introduction of the policy. The estimation shows that the state effect for the employment rate before the policy is around 2.932 and significant at 10%. Being a restaurant in New Jersey therefore increases the amount of employed workers by almost 3. The state effect for the employment rate after the policy is slightly higher with 3.424 and is signficant at 5%.

```
# https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.665.635&rep=rep1&type=pdf
data5 <- subset(data, !is.na(HRSOPEN) & !is.na(INCTIME) & !is.na(NREGS) &
    !is.na(prop))
prop <- matchit(STATE ~ HRSOPEN + INCTIME + NREGS + CHAIN + CO_OWNED +
    WAGE_ST + BONUS + MEALS + PSODA, data = data5, method = "nearest",
    distance = data5$prop, ratio = 1, replace = TRUE)
# it does not use the proposenty score 'distance =
# prs_df$pr_score' / distance ='glm'
summary(prop)
```

```
##
## Call:
## matchit(formula = STATE ~ HRSOPEN + INCTIME + NREGS + CHAIN +
##     CO_OWNED + WAGE_ST + BONUS + MEALS + PSODA, data = data5,
##     method = "nearest", distance = data5$prop, replace = TRUE,
```

```
##     ratio = 1)
##
## Summary of Balance for All Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance        1.4317        0.2527          1.2252     1.4589    0.3468
## HRSOPEN        14.4387       14.4141          0.0086     0.8886    0.0292
## INCTIME        17.6111       19.7188         -0.2057     0.5716    0.0324
## NREGS           3.6973        3.3750          0.2436     1.2675    0.0412
## CHAIN           2.0920        2.2500         -0.1457     0.8233    0.0395
## CO_OWNED        0.3410        0.3906         -0.1047          .    0.0496
## WAGE_ST         4.6129        4.6292         -0.0473     0.9458    0.0446
## BONUS           0.2490        0.3125         -0.1467          .    0.0635
## MEALS           1.8812        2.0469         -0.3038     1.7252    0.0552
## PSODA           1.0634        0.9719          1.0395     1.4855    0.2079
##          eCDF Max
## distance   0.6021
## HRSOPEN    0.0717
## INCTIME    0.0790
## NREGS      0.0894
## CHAIN      0.0776
## CO_OWNED   0.0496
## WAGE_ST    0.1139
## BONUS      0.0635
## MEALS      0.1521
## PSODA      0.6198
##
##
## Summary of Balance for Matched Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance        1.4317        1.3332          0.1024     1.5183    0.0216
## HRSOPEN        14.4387       12.9176          0.5317     1.0682    0.1208
## INCTIME        17.6111       21.0575         -0.3364     0.5481    0.0568
## NREGS           3.6973        3.5556          0.1072     1.1098    0.0225
## CHAIN           2.0920        2.1992         -0.0989     1.2529    0.0632
## CO_OWNED        0.3410        0.3027          0.0808          .    0.0383
## WAGE_ST         4.6129        4.6843         -0.2070     0.9953    0.0881
## BONUS           0.2490        0.2605         -0.0266          .    0.0115
## MEALS           1.8812        1.9847         -0.1897     3.6168    0.0779
## PSODA           1.0634        1.0420          0.2429     1.9483    0.0621
##          eCDF Max Std. Pair Dist.
## distance   0.1149          0.1444
## HRSOPEN    0.3027          1.1424
## INCTIME    0.1418          1.2512
## NREGS      0.0651          0.9586
## CHAIN      0.1149          0.9892
## CO_OWNED   0.0383          0.8567
## WAGE_ST    0.2682          1.0763
## BONUS      0.0115          0.8417
## MEALS      0.1686          0.6394
## PSODA      0.3793          0.6268
##
## Sample Sizes:
##               Control Treated
## All             64.       261
```

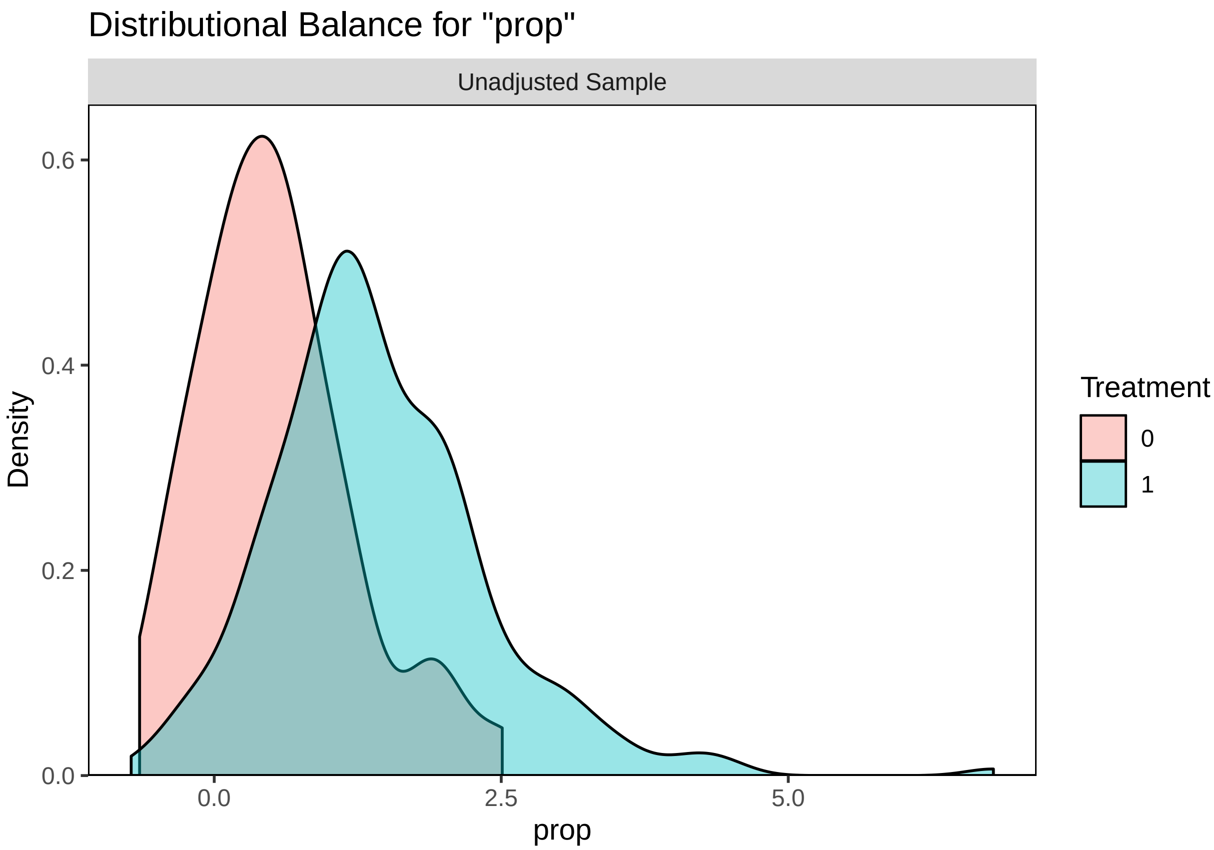
```
## Matched (ESS)   14.11     261
## Matched         42.       261
## Unmatched       22.         0
## Discarded        0.         0
```

```r
bal.tab(prop, disp.subclass = TRUE)
```

```
## Call
##  matchit(formula = STATE ~ HRSOPEN + INCTIME + NREGS + CHAIN +
##     CO_OWNED + WAGE_ST + BONUS + MEALS + PSODA, data = data5,
##     method = "nearest", distance = data5$prop, replace = TRUE,
##     ratio = 1)
##
## Balance Measures
##              Type Diff.Adj
## distance Distance   0.1024
## HRSOPEN    Contin.   0.5317
## INCTIME    Contin.  -0.3364
## NREGS      Contin.   0.1072
## CHAIN      Contin.  -0.0989
## CO_OWNED    Binary   0.0383
## WAGE_ST    Contin.  -0.2070
## BONUS       Binary  -0.0115
## MEALS      Contin.  -0.1897
## PSODA      Contin.   0.2429
##
## Sample sizes
##                      Control Treated
## All                    64.       261
## Matched (ESS)          14.11     261
## Matched (Unweighted)   42.       261
## Unmatched              22.         0
```

```r
data_m <- match.data(prop)
bal.plot(data_m, treat = data_m$STATE, var.name = "prop")
```

```
## Warning: Missing values exist in the covariates. Displayed values omit these
## observations.
```




```
weight = prop$weights
r1_m <- lm(EMP1 ~ STATE, data_m, weights = data_m$weights)
r2_m <- lm(EMP2 ~ STATE, data_m, weights = data_m$weights)
r1_m
```

```
##
## Call:
## lm(formula = EMP1 ~ STATE, data = data_m, weights = data_m$weights)
##
## Coefficients:
## (Intercept)        STATE
##      17.925        2.932
```

```
r2_m
```

```
##
## Call:
## lm(formula = EMP2 ~ STATE, data = data_m, weights = data_m$weights)
##
## Coefficients:
## (Intercept)        STATE
##      17.583        3.424
```

Q6

Next, we estimate the ATET on the change in employment in the restaurants. This basically equates to the difference between the two previous models in question 5. Now the average treatment effect on the treated decreases to 0.492. Furthermore, it is not significant.

```
# https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.665.635&rep=rep1&type=pdf
r3_m <- lm(difE ~ STATE, data_m, weights = data_m$weights)
stargazer(r1_m, r2_m, r3_m, column.labels = c(""), type = "text",
    title = "Treatment Effect", header = FALSE, label = "tab:reg1")
```

```
##
## Treatment Effect
## ==============================================================
##                                     Dependent variable:
##                                -------------------------------
##                                   EMP1       EMP2      difE
##
##                                   (1)        (2)       (3)
## --------------------------------------------------------------
## STATE                           2.932*     3.424**    0.492
##                                (1.588)    (1.394)    (1.363)
##
## Constant                       17.925***  17.583***  -0.342
##                                (1.474)    (1.294)    (1.265)
##
## --------------------------------------------------------------
## Observations                      303        303       303
## R2                               0.011      0.020     0.0004
## Adjusted R2                      0.008      0.016     -0.003
## Residual Std. Error (df = 301)   9.553      8.385      8.199
## F Statistic (df = 1; 301)        3.409*     6.034**    0.130
## ==============================================================
## Note:                           *p<0.1; **p<0.05; ***p<0.01
```

Q7 (Julian)

```
sum(data_m$prop * (((data_m$STATE * data_m$difE)/data_m$prop) - ((1 -
    data_m$STATE) * data_m$difE/(1 - data_m$prop))))/sum(data_m$prop)
```

```
## [1] 1.60952
```