Assignment

483670

September 16, 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 ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff 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
The mean number of employees in NJ are 'r tostring(NJemp1)' and "NJemp2" for the first and second wave i


'''r
NJemp1a <- mean(data$EMPFT[data$STATE == 1]) + mean(data$EMPPT[data$STATE ==
    1]) + mean(data$NMGRS[data$STATE == 1])
NJemp2a <- mean(data$EMPFT2[data$STATE == 1]) + mean(data$EMPPT2[data$STATE ==
    1]) + mean(data$NMGRS2[data$STATE == 1])
PAemp1a <- mean(data$EMPFT[data$STATE == 0]) + mean(data$EMPPT[data$STATE ==
    0]) + mean(data$NMGRS[data$STATE == 0])
PAemp2a <- mean(data$EMPFT2[data$STATE == 0]) + mean(data$EMPPT2[data$STATE ==
    0]) + mean(data$NMGRS2[data$STATE == 0])
EffNJa <- NJemp2a - NJemp1a
EffPAa <- PAemp2a - PAemp1a
data1 <- subset(data, STATUS2 == 1)
NJemp1b <- mean(data1$EMPFT[data$STATE == 1]) + mean(data1$EMPPT[data$STATE ==
    1]) + mean(data1$NMGRS[data$STATE == 1])
NJemp2b <- mean(data1$EMPFT2[data$STATE == 1]) + mean(data1$EMPPT2[data$STATE ==
    1]) + mean(data1$NMGRS2[data$STATE == 1])
PAemp1b <- mean(data1$EMPFT[data$STATE == 0]) + mean(data1$EMPPT[data$STATE ==
    0]) + mean(data1$NMGRS[data$STATE == 0])
PAemp2b <- mean(data1$EMPFT2[data$STATE == 0]) + mean(data1$EMPPT2[data$STATE ==
    0]) + mean(data1$NMGRS2[data$STATE == 0])
EffNJb <- NJemp2b - NJemp1b
EffPAb <- PAemp2b - PAemp1b
data <- data1
```

Q2

As we are looking at the difference in each shop we restrict the sample to those who responded in both waves. The effect of the minimum wage in the simple diff-in-diff is "dif_eff1". Once we add controls for the amount of time the shop is open and a proxy for its capacity (the number of registers) the magnitude of effect increases slightly to "dif_eff2". CAN WE CONTROL FOR THE NUMBER OF EMPLOYEES IMPACTED BY THE WAGE SHARE? OR IS THAT A MECHANISM??

```
data$EMP1 <- data$EMPFT + data$EMPPT + data$NMGRS
data$EMP2 <- data$EMPFT2 + data$EMPPT2 + data$NMGRS2
data$difE <- data$EMP1 - data$EMP2
m1 <- lm(difE ~ STATE, data)
m2 <- lm(difE ~ STATE + HRSOPEN + NREGS, data)
dif_eff1 <- m1$coefficients[2]
dif_eff2 <- m2$coefficients[2]
stargazer(m1, m2, column.labels = c("", "Controls"), type = "text",
    title = "Minimum Wage", header = FALSE, label = "tab:reg1")
```

```
##
```

```
## Minimum Wage
## ================================================================
##                                 Dependent variable:
##                     ---------------------------------------
##                                        difE
##                                               Controls
##                              (1)                 (2)
## ----------------------------------------------------------------
## STATE                      -2.327*             -2.391*
##                            (1.299)             (1.297)
##
## HRSOPEN                                        0.374**
##                                                (0.187)
##
## NREGS                                           0.088
##                                                (0.425)
##
## Constant                    1.838              -3.902
##                            (1.165)             (2.948)
##
## ----------------------------------------------------------------
## Observations                 378                 373
## R2                          0.008               0.021
## Adjusted R2                 0.006               0.014
## Residual Std. Error  10.022 (df = 376)     9.928 (df = 369)
## F Statistic         3.208* (df = 1; 376) 2.697** (df = 3; 369)
## ================================================================
## Note:                                *p<0.1; **p<0.05; ***p<0.01
```

Q3

The status variable provides problems to the command and has been left out of the analysis. It looks like there are potential differences between in the states on the number of employees. This difference seems to originate in a difference in the number of full-time employees in the first wave.

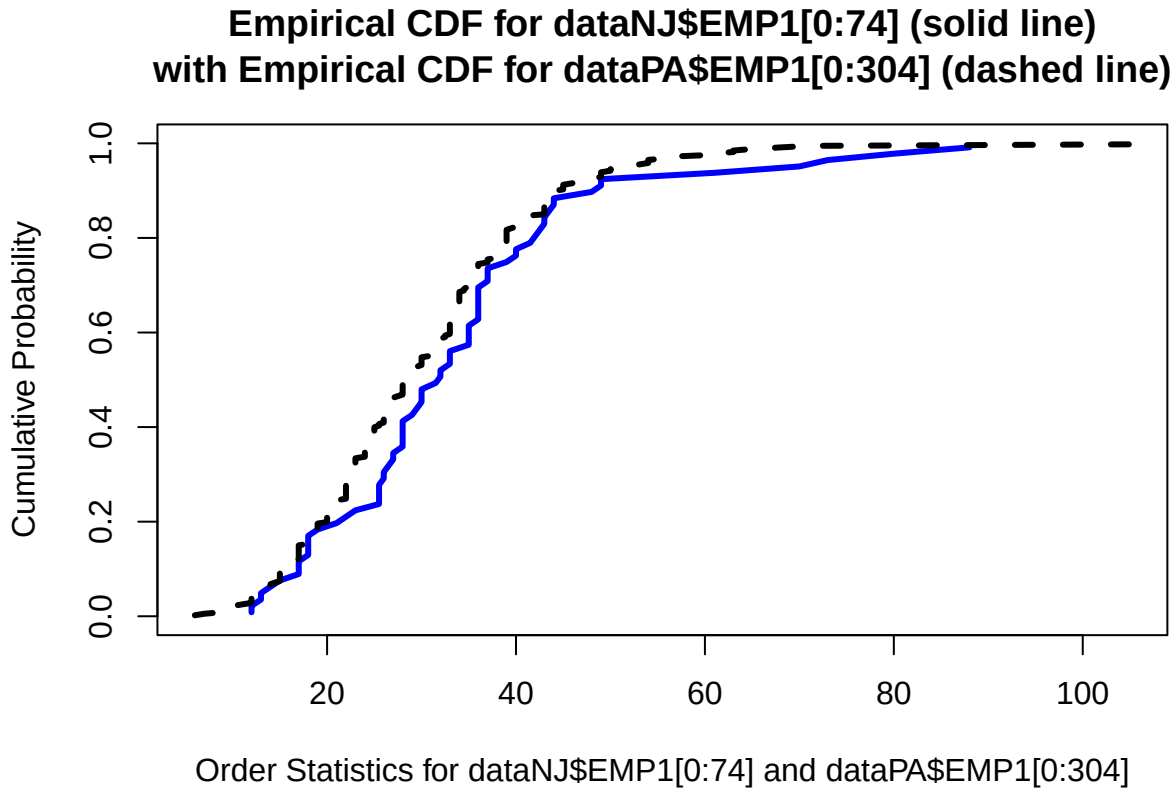
Comparing stepped dummy variables does not make an awful lot of sense.

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

Q4

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.

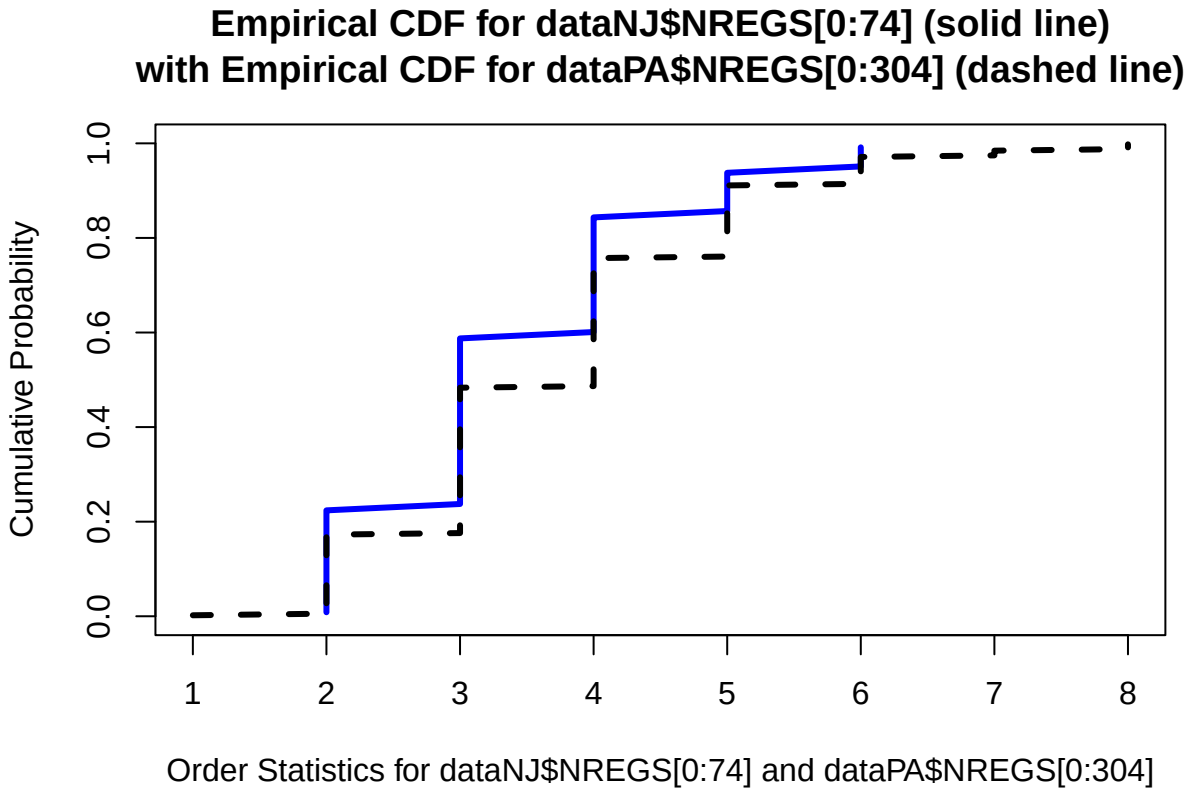
```
dataNJ <- subset(data, !STATE == 1)
dataPA <- subset(data, !STATE == 0)

gEMP1 <- cdfCompare(dataNJ$EMP1[0:74], dataPA$EMP1[0:304])
```



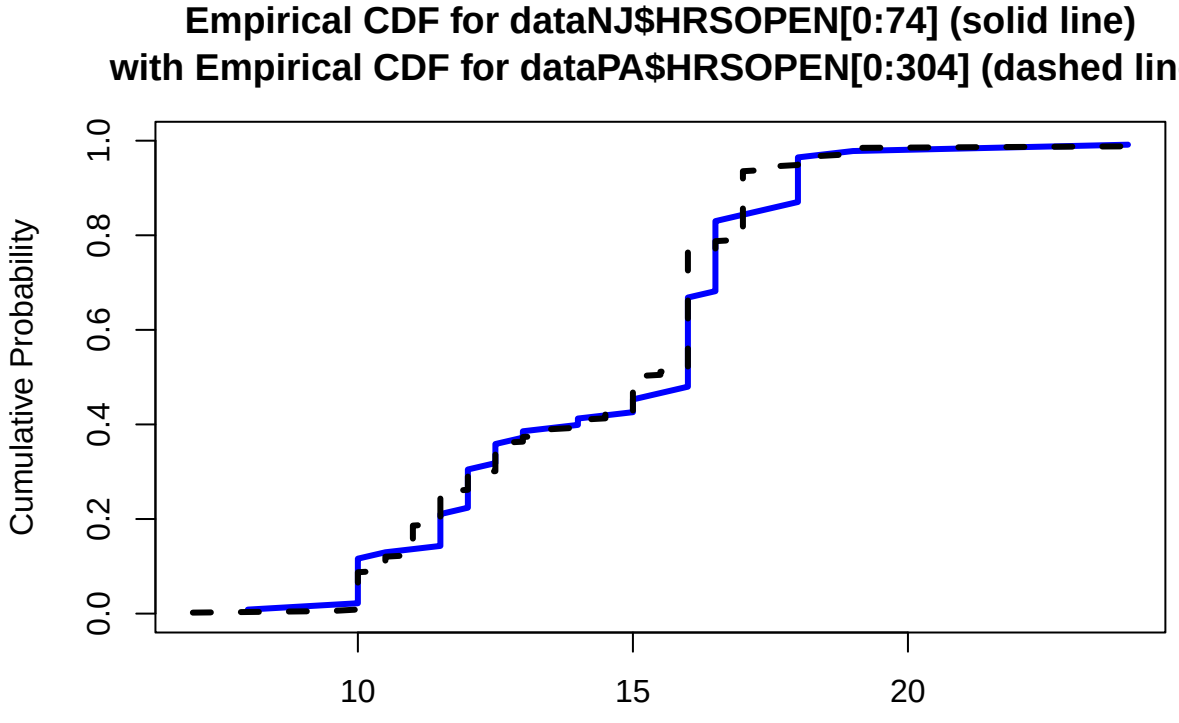

```
gNREGS <- cdfCompare(dataNJ$NREGS[0:74], dataPA$NREGS[0:304])
```

```
## Warning in is.not.finite.warning(y): There were 5 nonfinite values in y : 5 NA's
```

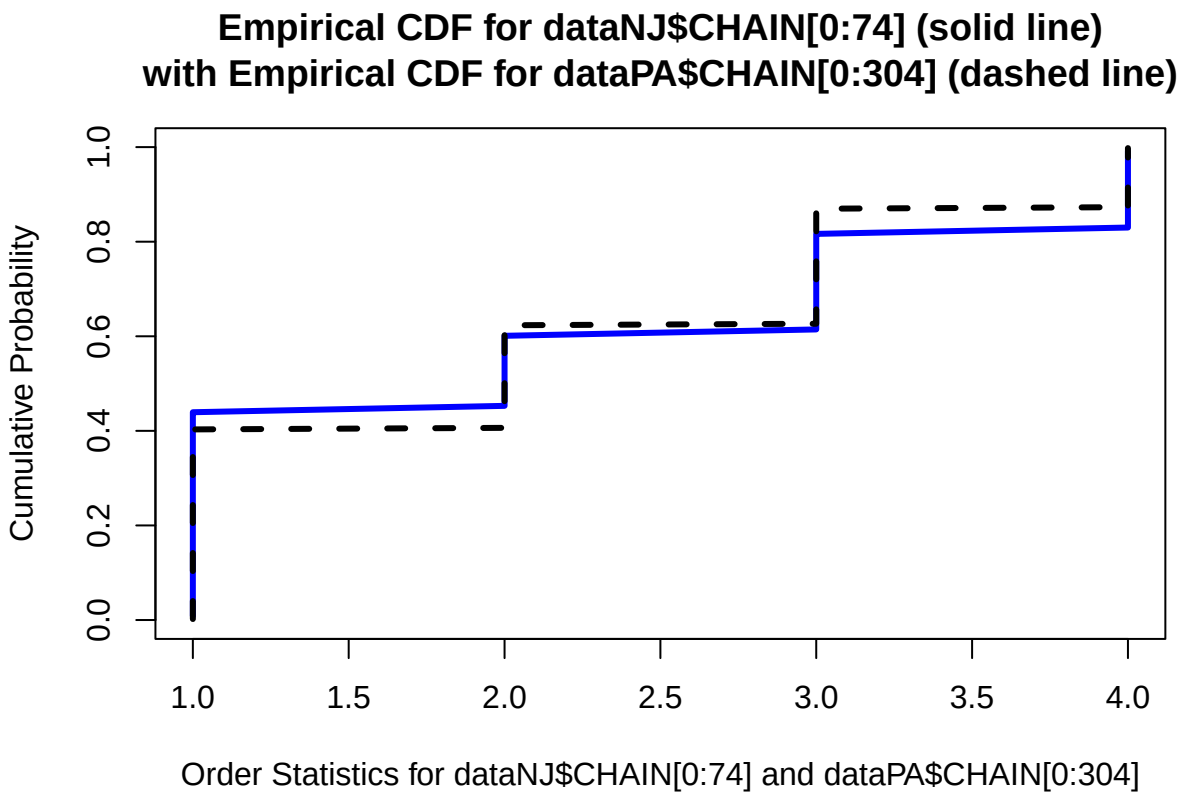
```
## Warning in cdfCompare(dataNJ$NREGS[0:74], dataPA$NREGS[0:304]): 5 observations
## with NA/NaN/Inf in 'y' removed.
```




```
gHRSOPEN <- cdfCompare(dataNJ$HRSOPEN[0:74], dataPA$HRSOPEN[0:304])
```

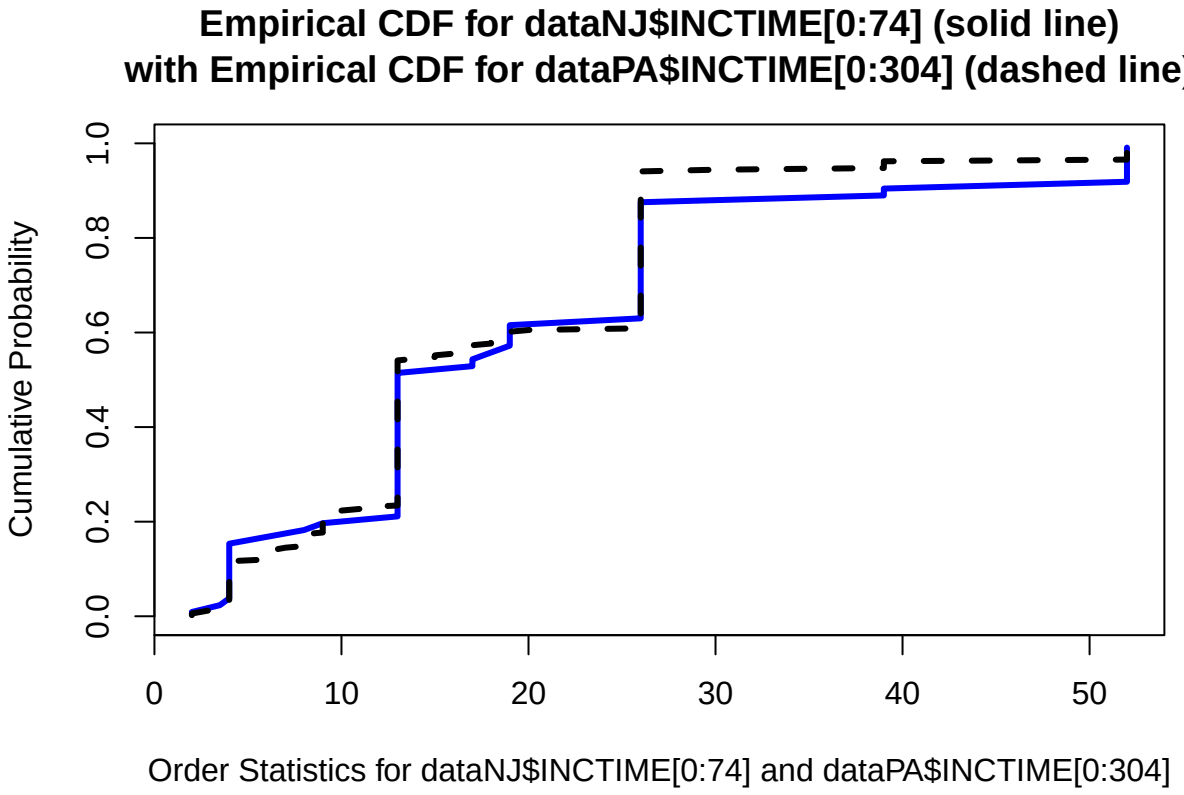


```
gCHAIN <- cdfCompare(dataNJ$CHAIN[0:74], dataPA$CHAIN[0:304])
```




```
gINCTIME <- cdfCompare(dataNJ$INCTIME[0:74], dataPA$INCTIME[0:304])
```

```
## Warning in is.not.finite.warning(x): There were 5 nonfinite values in x : 5 NA's

## Warning in cdfCompare(dataNJ$INCTIME[0:74], dataPA$INCTIME[0:304]): 5
## observations with NA/NaN/Inf in 'x' removed.

## Warning in is.not.finite.warning(y): There were 24 nonfinite values in y : 24
## NA's

## Warning in cdfCompare(dataNJ$INCTIME[0:74], dataPA$INCTIME[0:304]): 24
## observations with NA/NaN/Inf in 'y' removed.
```




```
# hEMP1 <- hist(dataNJ$EMPFT) hEMP2 <- hist(dataPA$EMPFT)

# plot(ha,col = 'blue' , xlim = c(0, 50), ylim = c(0,100)) plot
# (hb, col = 'red', add = TRUE)
probit <- glm(STATE ~ HRSOPEN + INCTIME + NREGS + CHAIN, family = binomial(link = "probit"),
    data = data)
data$prop <- predict.glm(probit, data)
stargazer(probit, column.labels = c("", ""), type = "text", title = "Minimum Wage",
    header = FALSE, label = "tab:probit")
```

```
##
## Minimum Wage
## ================================================
##                         Dependent variable:
##                     ---------------------------
##                                STATE
##
## ------------------------------------------------
## HRSOPEN                       -0.013
##                               (0.030)
##
## INCTIME                       -0.006
##                               (0.007)
##
## NREGS                         0.112*
##                               (0.066)
##
## CHAIN                         -0.040
```

```
##                               (0.071)
##
## Constant                     0.827*
##                               (0.502)
##
## ---------------------------------------------
## Observations                    345
## Log Likelihood               -170.480
## Akaike Inf. Crit.             350.960
## =============================================
## Note:             *p<0.1; **p<0.05; ***p<0.01
```

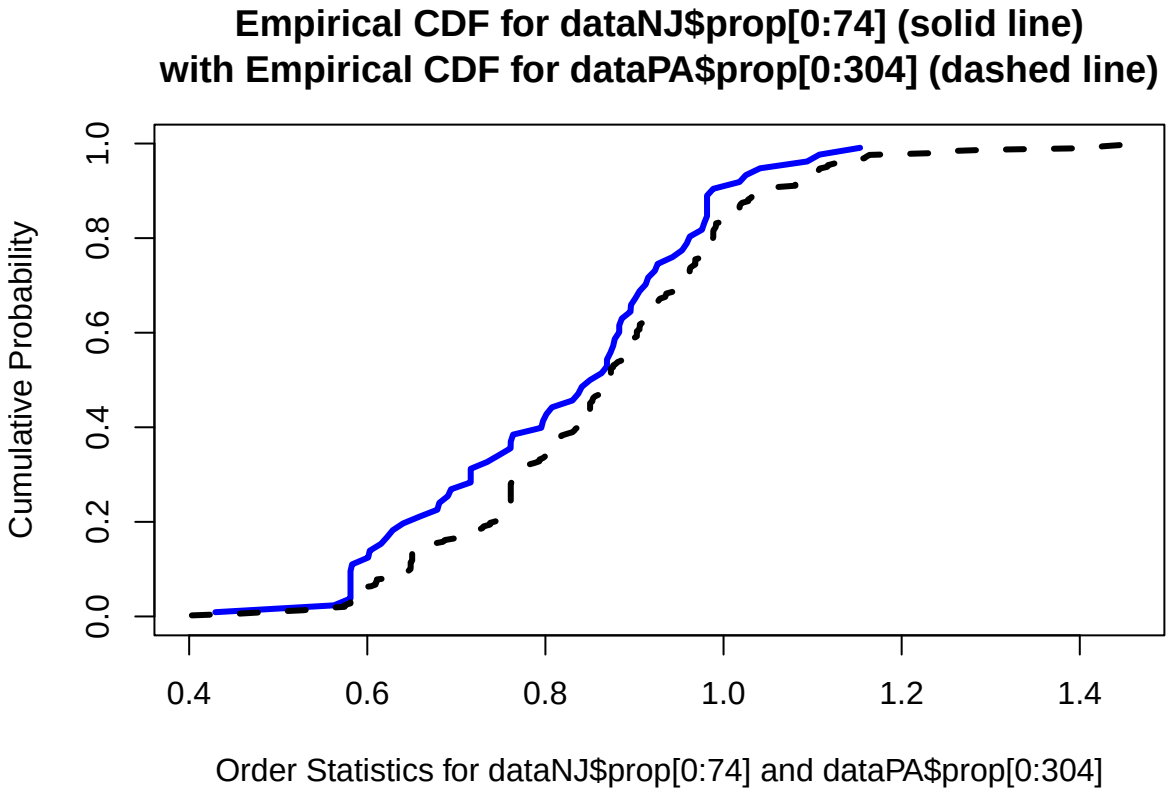
```
dataNJ <- subset(data, !STATE == 1)
dataPA <- subset(data, !STATE == 0)
graph1 <- cdfCompare(dataNJ$prop[0:74], dataPA$prop[0:304])
```

```
## Warning in is.not.finite.warning(x): There were 5 nonfinite values in x : 5 NA's
```

```
## Warning in cdfCompare(dataNJ$prop[0:74], dataPA$prop[0:304]): 5 observations
## with NA/NaN/Inf in 'x' removed.
```

```
## Warning in is.not.finite.warning(y): There were 28 nonfinite values in y : 28
## NA's
```

```
## Warning in cdfCompare(dataNJ$prop[0:74], dataPA$prop[0:304]): 28 observations
## with NA/NaN/Inf in 'y' removed.
```



Q5

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))
data5$Treat <- ifelse(data5$STATE == 1, 0, 1)
prop <- matchit(Treat ~ HRSOPEN + INCTIME + NREGS, data = data5, method = "nearest",
    distance = "glm", ratio = 1, replace = FALSE)

summary(prop)
```

```
##
## Call:
## matchit(formula = Treat ~ HRSOPEN + INCTIME + NREGS, data = data5,
##     method = "nearest", distance = "glm", replace = FALSE, ratio = 1)
##
## Summary of Balance for All Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance        0.2093        0.1977          0.2612     1.0998    0.0595
## HRSOPEN        14.4420       14.4982         -0.0185     1.1422    0.0312
## INCTIME        19.3696       17.8062          0.1179     1.5909    0.0282
## NREGS           3.3913        3.6848         -0.2573     0.7552    0.0376
##          eCDF Max
## distance   0.1341
## HRSOPEN    0.0761
## INCTIME    0.0652
## NREGS      0.0833
##
##
## Summary of Balance for Matched Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance        0.2093        0.2089          0.0075     1.0222    0.0035
## HRSOPEN        14.4420       14.4348          0.0024     1.2542    0.0277
## INCTIME        19.3696       20.2464         -0.0661     1.1474    0.0340
## NREGS           3.3913        3.4493         -0.0508     0.9110    0.0145
##          eCDF Max Std. Pair Dist.
## distance   0.0435          0.0155
## HRSOPEN    0.0580          0.7492
## INCTIME    0.0870          0.4684
## NREGS      0.0725          0.3050
##
## Sample Sizes:
##           Control Treated
## All           276      69
## Matched        69      69
## Unmatched     207       0
## Discarded       0       0
```

```
data_m <- match.data(prop)

r1_m <- lm(EMP1 ~ Treat + HRSOPEN + INCTIME + NREGS, data_m)
r2_m <- lm(EMP2 ~ Treat + HRSOPEN + INCTIME + NREGS, data_m)
```

Q6

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.665.635&rep=rep1&type=pdf

```
r3_m <- lm(difE ~ Treat + HRSOPEN + INCTIME + NREGS, data_m)
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)
## ---------------------------------------------------------------
## Treat                             2.673      1.278     1.395
##                                  (1.977)    (1.693)   (1.971)
##
## HRSOPEN                          2.753***   2.328***   0.425
##                                  (0.380)    (0.325)   (0.378)
##
## INCTIME                           -0.014     -0.015    0.001
##                                  (0.078)    (0.067)   (0.078)
##
## NREGS                           -2.883***  -2.562***  -0.322
##                                  (0.935)    (0.801)   (0.933)
##
## Constant                          0.991      5.384    -4.392
##                                  (5.504)    (4.712)   (5.487)
##
## ---------------------------------------------------------------
## Observations                       138        138       138
## R2                                0.292      0.282     0.013
## Adjusted R2                       0.271      0.260    -0.016
## Residual Std. Error (df = 133)   11.599     9.932     11.564
## F Statistic (df = 4; 133)      13.712***  13.060***   0.450
## ===============================================================
## Note:                              *p<0.1; **p<0.05; ***p<0.01
```