**2023 Causal Inference Workshop: Stata, R, and Python Sessions**

**Day 2. Implementing simple balancing strategies**

Underlying Paper: Black, Espin-Sanchez, French, and Litvak, *The Long-term Effect of Health Insurance on Near-Elderly Health and Mortality*, 3 **American Journal of Health Economics** 281-311 (2017)

**Answer Sheet**

**1) Question**: Use several different methods for creating a balanced sample: propensity score matching (PSM), nnmatch (nearest neighbor matching); and IPW. Then combine each with regression to estimate treatment effects over time since the initial measurement of insurance status in 1992. For the matching methods, use 1:1 matching. Use nnmatch without bias correction, which substitutes for regression; you can’t do both.

**STATA Code for Step 1**:

\***Locals for regressions** (these are the same local variables, for the same covariates, that we used for the day 1 solutions). The full paper includes many additional covariates, which lead to different results for later waves.

local two\_power\_age = "age c.age#c.age" //quadratic in age

local comorbidities = "i.diabetes\_wave\_1 i.cancer\_wave\_1 i.high\_blood\_pressure\_wave\_1 i.lung\_disease\_wave\_1 i.heart\_disease\_wave\_1 i.stroke\_wave\_1 i.psychiatric\_disease\_wave\_1 i.arthritis\_wave\_1 i.ulcer\_wave\_1 i.CESD\_wave\_1”

local money = "i.hh\_logincome\_quintile\_wave\_1 i.hh\_earning\_quintile\_wave\_1" //quintiles of logincome and earnings

local labor = "i.not\_in\_laborforce\_wave\_1 i.partly\_retired\_wave\_1 i.fully\_retired\_wave\_1 i.unemployed\_wave\_1 i.employed\_pt\_wave\_1 i.employed\_ft\_wave\_1 i.veteran\_wave\_1"

\***Replace age values with age-51.** This will produce a more reasonable coefficient for the constant term but won’t change anything else.

replace age=age-51

\***Set seed** (so results from the logit regression are exactly the same each time you run it)

set seed 12354

\*code is provided below for both the psmatch2 and teffects psmatch commands

\* **psmatch2 command**

psmatch2 no\_insurance\_wave\_1 `two\_power\_age' i.gender i.race `money' `labor' `comorbidities' `other' i.gender#i.race i.gender#c.(`two\_power\_age') i.race#c.(`two\_power\_age'), logit ai(1) outcome(death\_by\_wave\_2) n(1) ties

Note: The n(1) option specifies 1:1 matching. The ties option is one option for handling the situation where you have multiple, equally good control units. The ai option specifies Abadie-Imbens standard errors. For teffects psmatch, these are default options.

\***teffects psmatch for treated**

teffects psmatch (death\_by\_wave\_2) (no\_insurance\_wave\_1 `two\_power\_age' i.gender i.race `money' `labor' `comorbidities' `other' i.gender#i.race i.gender#c.(`two\_power\_age') i.race#c.(`two\_power\_age'), logit), atet

\***teffects nnmatch for treated**

teffects nnmatch (death\_by\_wave\_2 `two\_power\_age' i.gender i.race `money' `labor' `comorbidities' `other' i.gender#i.race i.gender#c.(`two\_power\_age') i.race#c.(`two\_power\_age')) (no\_insurance\_wave\_1 ), atet

\***teffects ipw for treated**

teffects ipw (death\_by\_wave\_2) (no\_insurance\_wave\_1 `two\_power\_age' i.gender i.race `money' `labor' `comorbidities' `other' i.gender#i.race i.gender#c.(`two\_power\_age') i.race#c.(`two\_power\_age')), atet

**2) Question**: Estimate the treatment effect (the mortality for the treated (uninsured) relative to the controls (insured) using each balancing method on the balanced sample, at each of waves 2-11.

**STATA Code for Step 2**:

set seed 12354

**\*Loop over the waves**

forval y=2/11{

**\*teffects psmatch for ATT**

teffects psmatch (death\_by\_wave\_`y') (no\_insurance\_wave\_1 `two\_power\_age' i.gender i.race `money' `labor' `comorbidities' `other' i.gender#i.race i.gender#c.(`two\_power\_age') i.race#c.(`two\_power\_age'), logit), atet vce(robust)

outreg2 using "$desktop\\ATT.xls", excel append ctitle("psmatch wave`y'") e(n1 vcetype cmd subcmd stat)

Note: The bdec, sdec, and tdec options export the coefficient, standard error, and t-statistics with the specified number of digits after the decimal point. However, with teffects, only coefficient and standard error can be exported and only with 4 digits.

**\*teffects nnmatch for ATT**

teffects nnmatch (death\_by\_wave\_`y' `two\_power\_age' i.gender i.race `money' `labor' `comorbidities' `other' i.gender#i.race i.gender#c.(`two\_power\_age') i.race#c.(`two\_power\_age')) (no\_insurance\_wave\_1), atet vce(robust)

outreg2 using "$desktop\\ATT.xls", excel append ctitle("nnmatch wave`y' e(n1 vcetype cmd subcmd stat)

**\*teffects ipw for ATT**

teffects ipw (death\_by\_wave\_`y') (no\_insurance\_wave\_1 `two\_power\_age' i.gender i.race `money' `labor' `comorbidities' `other' i.gender#i.race i.gender#c.(`two\_power\_age') i.race#c.(`two\_power\_age')), atet vce(robust)

outreg2 using "$desktop\\ATT.xls", excel append ctitle("ipw wave`y'") e(n1 vcetype cmd subcmd stat)

}

**Question 3**: Plot the results with each balancing approach: x-axis is waves; y-axis is treatment effect, show confidence intervals for each point estimate (in a way that you find visually appealing).

\*In order to plot the treatment effect coefficients in STATA:

\*1. Clean the IPW output from outreg2 over the 11 waves. For teffects IPW, outreg2 exports all the logit coefficients along with ATT and potential outcome means. Keep only ATT and standard errors.

\*2. For teffects nnmatch and teffects psmatch no cleaning action is required after outreg2 as the only coefficient exported is the ATT (and its standard error).

\*3. Remember to change the each outreg2 file format from “xls” to “xlsx”.

**\*Local for balancing methods.**

local methods = "psmatch nnmatch ipw"

**\*For each method, import wave 2-11 ATT.**

foreach k of local methods {

\*Load results.

import excel "$desktop\ATT\_`k'.xlsx", clear

\*Rename columns with wave number. Each column reports wave ATT and standard error.

rename (B C D E F G H I J K) (wave2 wave3 wave4 wave5 wave6 wave7 wave8 wave9 wave10 wave11)

**\*Drop lines and columns that are not needed from outreg2.** The rows report what we export with outreg2 under the option “e(n1 vcetype cmd subcmd stat)”

drop in 1/3

drop in 3/10

drop A

**\*This loop gets rid of the parenthesis the standard errors are reported with.**

forval y=2/11{

replace wave`y'=subinstr(wave`y', "(", "",.) if \_n==2

replace wave`y'=subinstr(wave`y', ")", "",.) if \_n==2

}

**\*Destring treatment coefficients and standard errors.**

destring wave\*, replace force

**\*Reshape data in long form so that each row reports wave [\*x] ATT coefficient and standard error.**

gen id=\_n

reshape long wave, i(id) j(wave\_number)

sort wave\_number id

gen std=.

by wave\_number id: replace std=wave if id==2

gsort wave\_number -id

carryforward std if std==., replace

drop if id==2

drop id

**\*Calculate Confidence interval.**

gen CI\_low=wave-1.96\*std

gen CI\_top=wave+1.96\*std

**\*Rename variables and assign labels.**

rename (wave) (ATET\_`k')

label var ATET\_`k' "ATT"

label var CI\_low "CI"

**\*Plot.**

twoway scatter ATET\_`k' wave\_number, connect(l) msymbol(none) lcol(purple) lwidth(medthick) /\*

\*/ || scatter CI\_low wave\_number, connect(l) lpattern(shortdash) lcol(red) lwidth(medthick) msymbol(none) /\*

\*/ || scatter CI\_top wave\_number, connect(l) lpatter(shortdash) lcol(red) lwidth(medthick) msymbol(none) /\*

\*/ yline(0, lcolor(black) lwidth(thin) lpattern(shortdash\_dot)) /\*

\*/ xtitle("Waves", size(medium)) /\*

\*/ xlabel(2 "2" 3 "3" 4 "4" 5 "5" 6 "6" 7 "7" 8 "8" 9 "9" 10 "10" 11 "11") /\*

\*/ legend(order(1 2) col(2) size(medium)) /\*

\*/ ytitle("Treatment effect", size(medium) height(5)) /\*

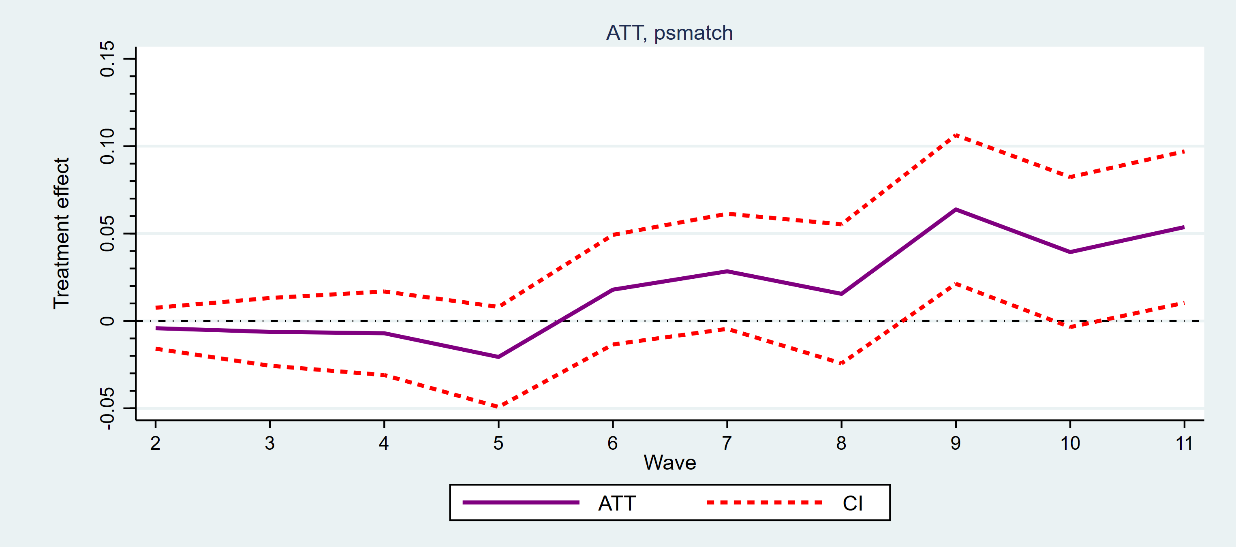
\*/ ylabel(-0.05 "-0.05" 0 "0" 0.05 "0.05" 0.1 "0.10" 0.15 "0.15", grid) /\*

\*/ title("ATT, `k'", size(medium)) subtitle("")

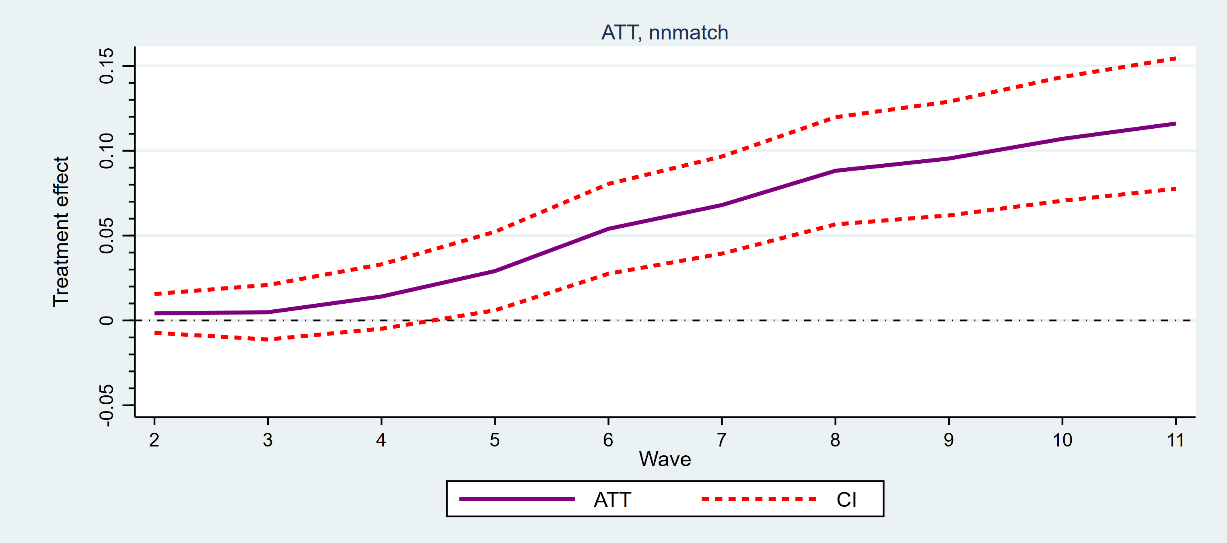
graph export "$desktop\`k'.png", as(png) replace width(3050) height(1350)

}

**Panel A**. teffects psmatch



**Panel B**. teffects nnmatch



**Panel C**. teffects ipw

