Stata Syntax for Section 5.8.6, Chapter 5

This section compares the propensity scores estimated by Rand-***gbm*** and Stata-***boost***

Stata Syntax to Run Stata-boost and Create Histograms and Boxplots

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

/\* Chapter 5 Section 5.8.6: Comparing Rand-gbm to Stata-boost

 In the file of g3aca1.dta, ps was created by R program Rand-gbm.

 The following syntax created ps1 using Stata user-developed program boost. \*/

cd "D:\psa\_e2\Chapter5\data"

use g3aca1.dta,replace

//The following syntax conducts a listwise deletion of missing data

quietly: regress intbl ageyc fmale blck whit hisp pcedu ipovl pcemft fthr dicsagg2 dicsint2 dccereg2 dccscom2 dccpros2 draggr2

set seed 1000

predict p

drop if p==.

set seed 1000

gen x=uniform()

sort x

//The following syntax runs Stata-boost (a user-created program)

program boost\_plugin, plugin using("D:\psa\_e2\Chapter5\data\boost64.dll")

boost intbl ageyc fmale blck whit hisp pcedu ipovl pcemft fthr dicsagg2 dicsint2 dccereg2 ///

 dccscom2 dccpros2 draggr2, distribution(logistic) maxiter(1000) trainfraction(0.8) ///

 pred(psb) inter(4) shrink(.0005)influence

sort id

sum ps psb

saveold "D:\psa\_e2\Chapter5\data\g3aca1\_.dta",replace

label var psb "Predicted Probability"

label var ps "Predicted Probability"

histogram psb if intbl==1, normal title(Treated) saving(g3aca1\_1,replace) xscale(range(0,1)) yscale(range(0,40))

histogram psb if intbl==0, normal title(Controls) saving(g3aca1\_0,replace) xscale(range(0,1)) yscale(range(0,40))

graph combine g3aca1\_0.gph g3aca1\_1.gph, title(Stata boost) xcommon ycommon saving(g3aca1\_boost.gph,replace)

histogram ps if intbl==1, normal title(Treated) saving(g3aca1\_1\_gbm,replace) xscale(range(0,1)) yscale(range(0,40))

histogram ps if intbl==0, normal title(Controls) saving(g3aca1\_0\_gbm,replace) xscale(range(0,1)) yscale(range(0,40))

graph combine g3aca1\_0\_gbm.gph g3aca1\_1\_gbm.gph, title(Rand gbm) xcommon ycommon saving(g3aca1\_gbm.gph,replace)

label define c 1 Treated 0 Control

label values intbl c

graph box psb, over(intbl) title(Stata boost) yscale(range(.2,.8))

save g1.gph,replace

graph box ps, over(intbl) title(Rand gbm) yscale(range(.2,.8))

save g2.gph,replace

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

R Syntax to Run Optimal Pair Matching Using the Two Propensity Scores

(Results are two ASCII files showing matched pairs: gbm\_pm.dat and boost\_pm.dat)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

#This run uses pscores generated by Rand-gbm (ps) and Stata-boost (psb)

#to compare pair matching results

set.seed(10)

setwd ("D:/psa\_e2/Chapter5/data")

library(foreign)

mc <- read.dta("g3aca1\_.dta")

library(optmatch)

#pairmatch pm using Rand-gbm

attach(mc)

prank <- rank(ps)

names(prank) <- id

d1 <- outer(prank[intbl==1], prank[intbl==0], "-")

d1 <- abs(d1)

pm <- pairmatch(d1, controls=1)

(pm.d <- matched.distances(pm,d1,pres=TRUE))

unlist(pm.d,max)

mean(unlist(pm.d))

sum(unlist(pm.d))

stratumStructure(pm)

write.table (pm, quote=FALSE, sep=",", col.names=FALSE, file="D:/psa\_e2/Chapter5/data/gbm\_pm.dat")

#pairmatch pm using Stata-boost

attach(mc)

prank <- rank(psb)

names(prank) <- id

d1 <- outer(prank[intbl==1], prank[intbl==0], "-")

d1 <- abs(d1)

pm1 <- pairmatch(d1, controls=1)

(pm1.d <- matched.distances(pm1,d1,pres=TRUE))

unlist(pm1.d,max)

mean(unlist(pm1.d))

sum(unlist(pm1.d))

stratumStructure(pm1)

write.table (pm1, quote=FALSE, sep=",", col.names=FALSE, file="D:/psa\_e2/Chapter5/data/boost\_pm.dat")

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Stata Syntax to Merge the Two Matched Sets to the Original Data Files, and Run Post-Matching Imbalance Check

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

/\* Chapter 5 Section 5.8.6

 After pair matching, we now merge the matched

 sets to the original data and perform imbalance

 check for both programs \*/

cd "D:\psa\_e2\Chapter5\data"

set more off

use g3aca1\_.dta,replace

sort id

save g3aca1\_,replace

clear

insheet id gbm\_pm using "D:\psa\_e2\Chapter5\data\gbm\_pm.dat"

sort id

merge id using g3aca1\_

save g3aca1\_a, replace

tab gbm\_pm intbl

use g3aca1\_a,replace

drop \_merge

sort id

save g3aca1\_a,replace

clear

insheet id b\_pm using "D:\psa\_e2\Chapter5\data\boost\_pm.dat"

sort id

merge id using g3aca1\_a

drop x

save g3aca1\_new, replace

tab b\_pm intbl

//imbalance check

imbalance g3aca1\_new ageyc intbl gbm\_pm ageyc\_gbm\_pm

imbalance g3aca1\_new fmale intbl gbm\_pm fmale\_gbm\_pm

imbalance g3aca1\_new blck intbl gbm\_pm blck\_gbm\_pm

imbalance g3aca1\_new whit intbl gbm\_pm whit\_gbm\_pm

imbalance g3aca1\_new hisp intbl gbm\_pm hisp\_gbm\_pm

imbalance g3aca1\_new pcedu intbl gbm\_pm pcedu\_gbm\_pm

imbalance g3aca1\_new ipovl intbl gbm\_pm ipovl\_gbm\_pm

imbalance g3aca1\_new pcemft intbl gbm\_pm pcemft\_gbm\_pm

imbalance g3aca1\_new fthr intbl gbm\_pm fthr\_gbm\_pm

imbalance g3aca1\_new ageyc intbl b\_pm ageyc\_b\_pm

imbalance g3aca1\_new fmale intbl b\_pm fmale\_b\_pm

imbalance g3aca1\_new blck intbl b\_pm blck\_b\_pm

imbalance g3aca1\_new whit intbl b\_pm whit\_b\_pm

imbalance g3aca1\_new hisp intbl b\_pm hisp\_b\_pm

imbalance g3aca1\_new pcedu intbl b\_pm pcedu\_b\_pm

imbalance g3aca1\_new ipovl intbl b\_pm ipovl\_b\_pm

imbalance g3aca1\_new pcemft intbl b\_pm pcemft\_b\_pm

imbalance g3aca1\_new fthr intbl b\_pm fthr\_b\_pm

clear

use ageyc\_gbm\_pm, replace

append using fmale\_gbm\_pm

append using blck\_gbm\_pm

append using whit\_gbm\_pm

append using hisp\_gbm\_pm

append using pcedu\_gbm\_pm

append using ipovl\_gbm\_pm

append using pcemft\_gbm\_pm

append using fthr\_gbm\_pm

append using ageyc\_b\_pm

append using fmale\_b\_pm

append using blck\_b\_pm

append using whit\_b\_pm

append using hisp\_b\_pm

append using pcedu\_b\_pm

append using ipovl\_b\_pm

append using pcemft\_b\_pm

append using fthr\_b\_pm

list

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Stata Syntax to Run Postmatching Regression Analysis Using Difference Scores

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

//Regressing difference-scores of Y on difference-scores of X

//gbm

use g3aca1\_new, replace

drop if intbl==0

drop if gbm\_pm=="NA"

sort gbm\_pm

rename icsacach y1

rename ageyc age1

rename fmale fmale1

rename blck blck1

rename whit whit1

rename hisp hisp1

rename pcedu pcedu1

rename ipovl ipovl1

rename pcemft pcemft1

rename fthr fthr1

keep id gbm\_pm y1 age1 fmale1 blck1 whit1 hisp1 pcedu1 ipovl1 pcemft1 fthr1 schbl

save "c:\tmp\y1", replace

use g3aca1\_new, replace

drop if intbl==1

drop if gbm\_pm=="NA"

sort gbm\_pm

rename icsacach y0

rename ageyc age0

rename fmale fmale0

rename blck blck0

rename whit whit0

rename hisp hisp0

rename pcedu pcedu0

rename ipovl ipovl0

rename pcemft pcemft0

rename fthr fthr0

keep id gbm\_pm y0 age0 fmale0 blck0 whit0 hisp0 pcedu0 ipovl0 pcemft0 fthr0 schbl

merge gbm\_pm using "c:\tmp\y1"

gen y=y1-y0

gen age=age1-age0

gen fmale=fmale1-fmale0

gen blck=blck1-blck0

gen whit=whit1-whit0

gen hisp=hisp1-hisp0

gen pcedu=pcedu1-pcedu0

gen ipovl=ipovl1-ipovl0

gen pcemft=pcemft1-pcemft0

gen fthr=fthr1-fthr0

save reg\_d\_gbm, replace

//boost

use g3aca1\_new, replace

drop if intbl==0

drop if b\_pm=="NA"

sort b\_pm

rename icsacach y1

rename ageyc age1

rename fmale fmale1

rename blck blck1

rename whit whit1

rename hisp hisp1

rename pcedu pcedu1

rename ipovl ipovl1

rename pcemft pcemft1

rename fthr fthr1

keep id b\_pm y1 age1 fmale1 blck1 whit1 hisp1 pcedu1 ipovl1 pcemft1 fthr1 schbl

save "c:\tmp\y1", replace

use g3aca1\_new, replace

drop if intbl==1

drop if b\_pm=="NA"

sort b\_pm

rename icsacach y0

rename ageyc age0

rename fmale fmale0

rename blck blck0

rename whit whit0

rename hisp hisp0

rename pcedu pcedu0

rename ipovl ipovl0

rename pcemft pcemft0

rename fthr fthr0

keep id b\_pm y0 age0 fmale0 blck0 whit0 hisp0 pcedu0 ipovl0 pcemft0 fthr0 schbl

merge b\_pm using "c:\tmp\y1"

gen y=y1-y0

gen age=age1-age0

gen fmale=fmale1-fmale0

gen blck=blck1-blck0

gen whit=whit1-whit0

gen hisp=hisp1-hisp0

gen pcedu=pcedu1-pcedu0

gen ipovl=ipovl1-ipovl0

gen pcemft=pcemft1-pcemft0

gen fthr=fthr1-fthr0

save reg\_d\_boost, replace

use reg\_d\_gbm, replace

regress y age fmale blck whit hisp pcedu ipovl pcemft fthr, robust cluster(schbl)

use reg\_d\_boost, replace

regress y age fmale blck whit hisp pcedu ipovl pcemft fthr, robust cluster(schbl)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_