Econ 675 Assignment 3

Nathan Mather*

December 16, 2018

Contents

1 Question 1: Estimating Equations

1.1 Q1 Part 1

To show that these are valid moment conditions we just need to show that they are all equal to zero. We start with the IPW condition

$$\mathrm{E}\left[\psi_{IPW}(\boldsymbol{Z}_i;\theta_t(g))\right] = \mathrm{E}\left[\frac{D_i(t)\cdot g(Y_i(t))}{p_t(\boldsymbol{X}_i)} - \theta(g)\right] = \mathrm{E}\left[\mathrm{E}\left[\frac{D_i(t)\cdot g(Y_i(t))}{p_t(\boldsymbol{X}_i)}|\boldsymbol{X}_i\right]\right] - \theta(g)$$

$$= \mathrm{E}\left[\frac{1}{p_t(\boldsymbol{X}_i)}\mathrm{E}\left[D_i(t)\cdot g(Y_i(t))|\boldsymbol{X}_i\right]\right] - \theta(g)$$

Now notice that

$$\mathrm{E}[D_i(t)|\boldsymbol{X}_i] = Pr[D_i(t)=1|\boldsymbol{X}_i] = Pr[T_i=t|\boldsymbol{X}_i] = p_t(\boldsymbol{X}_i)$$

using this we get

$$\mathrm{E}\left[\psi_{IPW}(\boldsymbol{Z}_i;\theta_t(g))\right] = \mathrm{E}\left[\mathrm{E}\left[g(Y_i(t))|\boldsymbol{X}_i\right]\right] - \theta(g) = \mathrm{E}\left[g(Y_i(t))\right] - \theta(g) = 0$$

*Shouts out to Ani for the help with this. Could not have done it without you!

Next we check $\psi_{RI1,t}$

$$\mathrm{E}[\psi_{RI1,t}(\boldsymbol{Z}_i;\theta_t(g))] = \mathrm{E}[e_t(g;\boldsymbol{X}_i)] - \theta_t(g) = \mathrm{E}[\mathrm{E}[g(Y_i(t))|\boldsymbol{X}_i]] - \theta_t(g) = \mathrm{E}[g(Y_i(t))] - \theta_t(g) = 0$$

Next check $\psi_{RI2,t}$

$$\mathrm{E}[\psi_{RI2,t}(\boldsymbol{Z}_i;\theta_t(g))] = \mathrm{E}\left[\frac{D_i(t)\cdot e_t(g;\boldsymbol{X_i})}{p_t(\boldsymbol{X}_i)}\right] - \theta(g) = \mathrm{E}\left[\mathrm{E}\left[\frac{D_i(t)\cdot e_t(g;\boldsymbol{X_i})}{p_t(\boldsymbol{X}_i)}|\boldsymbol{X}_i\right]\right] - \theta(g) = \mathrm{E}[e_t(g;\boldsymbol{X}_i)] - \theta_t(g) = 0$$

Finally we check $\psi_{DR,t}$

$$\mathrm{E}[\psi_{DR,t}(\boldsymbol{Z}_i;\theta_t(g))] = \mathrm{E}\left[\frac{D_i(t)\cdot g(Y_i(t))}{p_t(\boldsymbol{X}_i)} - \theta(g)\right] - \mathrm{E}\left[\frac{e_t(g;\boldsymbol{X}_i)}{p_t(\boldsymbol{X}_i)}(D_i(t) - p_t(\boldsymbol{X}_i))\right]$$

This first terms are identical to the IPW condition so we need only check the following.

$$\mathrm{E}\left[\frac{e_t(g;\boldsymbol{X}_i)}{p_t(\boldsymbol{X}_i)}(D_i(t) - p_t(\boldsymbol{X}_i))\right] = \mathrm{E}\left[\frac{e_t(g;\boldsymbol{X}_i)D_i(t)}{p_t(\boldsymbol{X}_i)} - pe_t(g;\boldsymbol{X}_i)\right] = \theta_t(g) - \theta_t(g) = 0$$

So all functions are valid moment conditions

1.2 Q1 Part 2

The plug-in IPW estimator is

$$\hat{\theta}_{\texttt{IPW},t}(g) = \frac{1}{n}\sum_{i=1}^{n}\frac{D_i(t)g(Y_i)}{\hat{p}_t(\boldsymbol{X}_i)}$$

$\hat{p}_t(\boldsymbol{X}_i)$ is the estimated propensity score. Because this has multiple treatment levels we can estimate the propensity score with any suitable discrete choice model. For example the multinomial logit model.

The RD1 estimator is

$$\begin{aligned}\hat{\theta}_{\texttt{RI1},t}(g) = \hat{\mathrm{E}}[e_t(g;\boldsymbol{X}_i)] = \frac{1}{n}\sum_{i=1}^{n}\hat{\mathrm{E}}[g(Y_i(t))|\boldsymbol{X}_i] &= \frac{1}{n}\sum_{i=1}^{n}\hat{\mathrm{E}}[g(Y_i(t))|\boldsymbol{X}_i, D_i(t)=1] \\ &= \frac{1}{n}\sum_{i=1}^{n}\hat{\mathrm{E}}[g(Y_i)|\boldsymbol{X}_i, D_i(t)=1],\end{aligned}$$

where the second last equality uses the ignorability assumption. We just need to decide how to estimate this last term. We could probably use NLS or some nonparametric methid.
The plug-in 'hybrid' imputation estimator is

$$\hat{\theta}_{\texttt{RI2},t}(g) = \frac{1}{n}\sum_{i=1}^{n}\frac{D_i(t)\widehat{\mu}_t(\boldsymbol{X}_i)}{\hat{p}_t(\boldsymbol{X}_i)}.$$

Finally, the plug-in doubly robust estimator is given by

$$\begin{aligned}\hat{\theta}_{\texttt{DR},t}(g) &= \frac{1}{n}\sum_{i=1}^{n}\frac{D_i(t)g(Y_i)}{\hat{p}_t(\boldsymbol{X}_i)} - \frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{\mu}_t(\boldsymbol{X}_i)}{\hat{p}_t(\boldsymbol{X}_i)}(D_i(t) - \hat{p}_t(\boldsymbol{X}_i)) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i(t)(g(Y_i) - \widehat{\mu}_t(\boldsymbol{X}_i))}{\hat{p}_t(\boldsymbol{X}_i)} + \widehat{\mu}_t(\boldsymbol{X}_i)\right).\end{aligned}$$

As discussed in Abadie and Catteneo (2018), the relative performance of the above estimators depends on the features of the data generating process. In finite samples, IPW estimators become unstable when the propensity score approaches zero or one and regression imputation estimators may suffer from extrapolation biases. Doubly robust estimators include safeguards against bias caused by misspecification but impose additional specification choices that may affect the resulting estimate.

1.3 Q1 Part 3

Note that

$$\sigma_t^2 = \mathrm{V}[Y_i(t)] = \mathrm{E}\left[Y_i(t) - \mathrm{E}[Y_i(t)]\right]^2$$

Thus, we can estimate σ_t^2 using any of the Methods from 1.2, with $g(Y_i(t)) = \mathrm{E}\left[Y_i(t) - \mathrm{E}[Y_i(t)]\right]^2$. This would be a two-step estimator, since we would need to estimate $\mathrm{E}[Y_i(t)]$. To conduct the hypothesis test of $H_0 : \sigma_t^2 = \sigma^2 \ \forall t \in \mathcal{T}$ we would need to use an appropriate joint hypothesis testing procedure. One way to proceed would be test $H_0 : \sigma_t^2 - \sigma^2 = 0 \ \forall t \in \mathcal{T}$ and construct the vector $\widehat{\boldsymbol{\theta}} = (\hat{\sigma}_1^2 - \sigma^2, ..., \hat{\sigma}_T^2 - \sigma^2)'$, and then show $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to \mathcal{N}(0, V)$. Then, the Delta method implies $\sqrt{n}(||\widehat{\boldsymbol{\theta}}||^2 - ||\boldsymbol{\theta}_0||^2) \to \mathcal{N}(0, 4\boldsymbol{\theta}_0' V \boldsymbol{\theta}_0)$. Note that under the null $\boldsymbol{\theta}_0 = 0$, so we can now conduct the hypothesis test $H_0 : \boldsymbol{\theta}_0 = 0$ in the usual way, using an estimator for the asymptotic variance.

1.4 Q1 Part 4

No Thanks

2 Question 2: Estimating Average Treatment Effects

A few things didn't run in R but it all went through in STATA. Results are below. I only did one table because making it is tedious but the code for both programs is in the appendix

ATE

statistic	specificaiton	estimate_exp	std.error_exp	CI_L	CI_U	estimate_PSID	std.error_PSID	CI_L	CI_U
Mean Diff		1794	670	479	3109	-15204	656	-16490	-13919
OLS	a	1582	659	291	2873	6302	1209	3932	8673
OLS	b	1507	657	219	2795	4699	1027	2686	6712
OLS	c	1501	663	202	2800	4284	1031	2263	6306
Reg. Impute	a	1462	630	228	2697	-11195	1741	-14608	-7782
Reg. Impute	b	1454	631	218	2690	-10398	3549	-17355	-3442
Reg. Impute	c	1428	642	170	2685	-11920	3498	-18776	-5065
IPW	a	1537	630	303	2772	-13507	2800	-18996	-8019
IPW	b	1470	631	234	2706	-7246	3550	-14204	-288
IPW	c	1468	642	210	2726	-7487	3499	-14344	-629
D. Robust	a	1473	630	239	2707	-13507	2800	-18996	-8019
D. Robust	b	1451	631	215	2687	-11419	3549	-18376	-4463
D. Robust	c	1423	642	166	2682	-12504	3498	-19360	-5649
N1 Match	a	1829	780	302	3358	-15619	1153	-17880	-13359
N1 Match	b	1876	735	435	3316	-9350	3975	-17140	-1559
N1 Match	c	1672	726	248	3095	-9560	4034	-17467	-1656
P Match	a	1542	646	275	2808	-15859	6750	-29089	-2629
P Match	b	1489	765	-12	2989	8646	15056	-20863	38156
P Match	c	1257	677	-70	2584	-9562	4034	-17468	1657

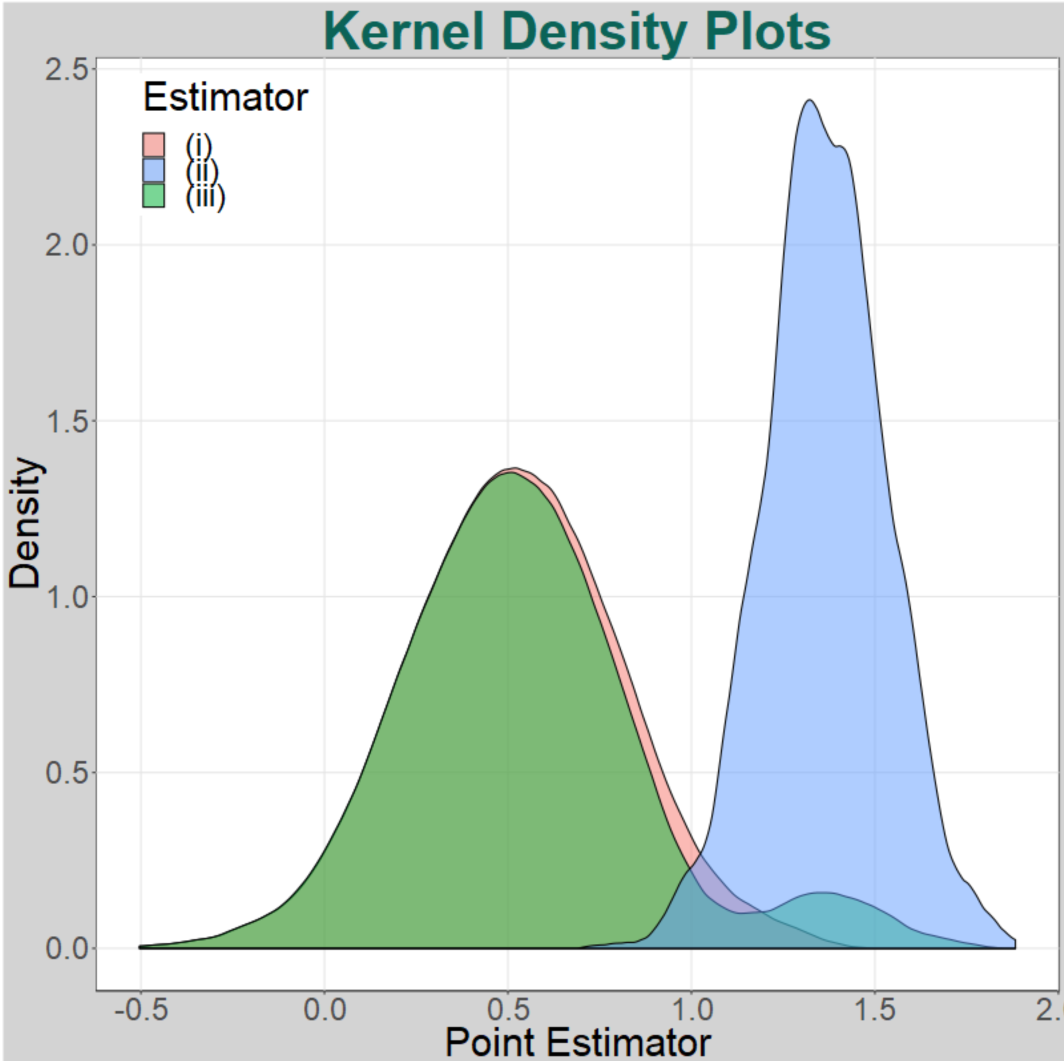
ATT

statistic	specificaiton	estimate_exp	std.error_exp	CI_L	CI_U	estimate_PSID	std.error_PSID	CI_L	CI_U
Mean Diff		1794	670	479	3109	-15204	656	-16490	-13919
OLS	a	1582	659	291	2873	6302	1209	3932	8673
OLS	b	1507	657	219	2795	4699	1027	2686	6712
OLS	c	1501	663	202	2800	4284	1031	2263	6306
Reg. Impute	a	1462	630	228	2697	-11195	1741	-14608	-7782
Reg. Impute	b	1454	631	218	2690	-10398	3549	-17355	-3442
Reg. Impute	c	1428	642	170	2685	-11920	3498	-18776	-5065
IPW	a	1537	630	303	2772	-13507	2800	-18996	-8019
IPW	b	1470	631	234	2706	-7246	3550	-14204	-288
IPW	c	1468	642	210	2726	-7487	3499	-14344	-629
D. Robust	a	1473	630	239	2707	-13507	2800	-18996	-8019
D. Robust	b	1451	631	215	2687	-11419	3549	-18376	-4463
D. Robust	c	1423	642	166	2682	-12504	3498	-19360	-5649
N1 Match	a	1829	780	302	3358	-15619	1153	-17880	-13359
N1 Match	b	1876	735	435	3316	-9350	3975	-17140	-1559
N1 Match	c	1672	726	248	3095	-9560	4034	-17467	-1656
P Match	a	1542	646	275	2808	-15859	6750	-29089	-2629
P Match	b	1489	765	-12	2989	8646	15056	-20863	38156
P Match	c	1257	677	-70	2584	-9562	4034	-17468	1657

3 Question 3: Post-model Seelction Inference

Summary Stats

estimator	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
i	-0.50	0.32	0.52	0.51	0.71	1.32
ii	0.77	1.26	1.36	1.36	1.47	1.88
iii	-0.50	0.32	0.52	0.54	0.72	1.66



3.1 Q3 Part 2

Coverage Rates

estimator	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
i	-0.50	0.32	0.52	0.51	0.71	1.32
ii	0.77	1.26	1.36	1.36	1.47	1.88
iii	-0.50	0.32	0.52	0.54	0.72	1.66

4 Appendix

4.1 R Code

pset 4 675

```r
#==========================#
# ==== Metrics 675 ps 4 ====
#==========================#


#=========================================#
# ==== Load packages, clear workspace ====
#=========================================#


  library(foreach)
  library(data.table)
  library(Matrix)
  library(ggplot2)
  library(sandwich)
  library(xtable)
  library(boot)
  library(CausalGAM)
  library(Hmisc)
library(mvtnorm)


  rm(list = ls(pos = ".GlobalEnv"), pos = ".GlobalEnv")
  options(scipen = 999)
  cat("\f")

#=====================================================#
# ==== Input data, add covariates and subset data ====
#=====================================================#

  lal_dt <- fread('C://Users/Nmath_000/Documents/MI_school/Second Year/675 Applied Econometrics/hw/hw4/l

  lal_dt[,log.re74 := log(re74+1)]
  lal_dt[,log.re75 := log(re75+1)]
  lal_dt[,age.sq := age^2]
  lal_dt[,educ.sq := educ^2]
  lal_dt[,age.cu := age^3]
  lal_dt[,black.u74 := black*u74]
  lal_dt[,educ.logre74 := educ*log.re74]

  # subset lal_dt for LaLonde control only
  lal_c1 <-  lal_dt[treat==1 | treat==0]

  # subset lal_dt for PSID control only
  lal_c2 <- lal_dt[treat==1 | treat==2]


  # Recode treatment indicate in PSID control dataset (recode 2's as 0's)
  lal_c2[,treat:=as.numeric(treat==1)]
```

```r
#=================================#
# ==== Create covariate lists ====
#=================================#


  z.a <- c("age", "educ", "black", "hisp", "married", "nodegr", "log.re74", "log.re75")
  z.b <- c(z.a,"age.sq","educ.sq", "u74", "u75" )
  z.c <- c(z.b, "age.cu", "black.u74", "educ.logre74")

  # make sure I didn't mis-type any (sohuld be empty sets)
  setdiff(z.a, colnames(lal_c1))
  setdiff(z.b, colnames(lal_c1))
  setdiff(z.c, colnames(lal_c1))


#==================================#
# ==== [1] Difference in means ====
#==================================#

  # run diff in diff
  dmeans.ll <- lm(re78~treat, data = lal_c1)
  dmeans.ps <- lm(re78~treat, data = lal_c2)

  # get robust se
  dmeans.ll <- data.table(tidy(coeftest(dmeans.ll, vcov = vcovHC(dmeans.ll, type = "HC1"))))
  dmeans.ps<- data.table(tidy(coeftest(dmeans.ps, vcov = vcovHC(dmeans.ps, type = "HC1"))))

  # just keep treat term
  dmeans.ll <- dmeans.ll[term == "treat", -c("term")]
  dmeans.ps <- dmeans.ps[term == "treat", -c("term")]

  dmeans.ll <- round(dmeans.ll, 2)
  dmeans.ps <- round(dmeans.ps, 2)

  # Compute 95% CIs
  dmeans.ll[, CI_lower := estimate - 1.96*std.error]
  dmeans.ll[, CI_upper := estimate + 1.96*std.error]
  dmeans.ll[, CI := paste0("(", CI_lower, ", ", CI_upper, ")")]

  dmeans.ps[, CI_lower := estimate - 1.96*std.error]
  dmeans.ps[, CI_upper := estimate + 1.96*std.error]
  dmeans.ps[, CI := paste0("(", CI_lower, ", ", CI_upper, ")")]


  # keep what I need and put them in same table
  dmeans.ll <- dmeans.ll[, c("estimate", "std.error", "CI" )]
  dmeans.ps <- dmeans.ps[, c("estimate", "std.error", "CI" )]

  setnames(dmeans.ll, colnames(dmeans.ll), paste0(colnames(dmeans.ll), '_exp'))
  setnames(dmeans.ps, colnames(dmeans.ps), paste0(colnames(dmeans.ps), '_PSID'))

  out_dmeans <- cbind(dmeans.ll, dmeans.ps)
```

```r
#==================#
# ==== [2] OLS ====
#==================#


  # write wrapper function to get what I need
  # # define variables for line by line debug
  # in_dt=lal_c1
  # in_z = z.a
  # in_spec = "a"
  # in_c_label = "exp"

  # start function wrapper
  ols_wrap <- function(in_dt, in_z, in_spec, in_c_label ){

    # run ols
    reg_form <- reg_form <- as.formula(paste("re78~", paste(c("treat", in_z), collapse="+")))
    ols_out <- lm(reg_form, data = in_dt)

    # get robust se
    ols_out <- data.table(tidy(coeftest(ols_out, vcov = vcovHC(ols_out, type = "HC1"))))

    # grab the term we care about
    ols_out <- ols_out[term == "treat", -c("term")]

    # round
    ols_out <- round(ols_out, 2)

    # do CI
    ols_out[, CI_lower := estimate - 1.96*std.error]
    ols_out[, CI_upper := estimate + 1.96*std.error]
    ols_out[, CI := paste0("(", CI_lower, ", ", CI_upper, ")")]
    ols_out <- ols_out[, c("estimate", "std.error", "CI" )]

    # put label in
    setnames(ols_out, colnames(ols_out), paste0(colnames(ols_out),"_", in_c_label )  )
    ols_out[, specification := in_spec]
    return(ols_out[])
  }

  ols_wrap(in_z = z.a, in_spec = "a" ,in_dt=lal_c1, in_c_label = "exp" )

  # run on all exprimental data
  exp_ols <- mapply(ols_wrap,in_z = list(z.a, z.b, z.c), in_spec = list("a", "b", "c") ,in_dt= list(lal
  exp_ols <- rbindlist(exp_ols)

  # run on psid
  psid_ols <- mapply(ols_wrap, in_z = list(z.a, z.b, z.c), in_spec = list("a", "b", "c") ,in_dt= list(la
  psid_ols <- rbindlist(psid_ols)

  # put it in one table
  ols_out <- merge(psid_ols, exp_ols, by = "specification")
```

```
# cant get the standard errors this way. Gonna use a package to do it below
# #=================================#
# # ==== Regression Imputation ====
# #=================================#
#
#
#   # start function wrapper
#   in_dt = lal_dt
#   in_z = z.a
#   in_spec = "a"
#   reg_imp <- function(in_dt, in_z, in_spec ){
#
#     # run ols reg on each treatment group
#     reg_form <- reg_form <- as.formula(paste("re78~", paste(in_z, collapse="+")))
#     ols.treat      <- lm(reg_form, data = in_dt[treat == 1])
#     ols.control.ll  <- lm(reg_form, data = in_dt[treat == 0])
#     ols.control.ps  <- lm(reg_form, data = in_dt[treat == 2])
#
#     # make matrix of data so I can calculate ATE
#     # y values
#     Y.treat        = lal_dt[treat==1, "re78"]
#     Y.control.ll   = lal_dt[treat==0, "re78"]
#     Y.control.ps   = lal_dt[treat==2, "re78"]
#
#     # x values
#     X.treat <- data.table(const =1 , lal_dt[treat==1, in_z, with = FALSE])
#     X.control.ll <- data.table(const =0 , lal_dt[treat==1, in_z, with = FALSE])
#     X.control.ps <- data.table(const =2 , lal_dt[treat==1, in_z, with = FALSE])
#
#
#     # Impute `individual treatment effects`
#     tvec.ri.treat.ll      = as.matrix(X.treat)%*%(as.vector(ols.treat$coefficients)-as.vector(ols.con
#     tvec.ri.treat.ps      = as.matrix(X.treat)%*%(as.vector(ols.treat$coefficients)-as.vector(ols.con
#
#     tvec.ri.control.ll    = as.matrix(X.control.ll)%*%(as.vector(ols.treat$coefficients)-as.vector(ol
#     tvec.ri.control.ps    = as.matrix(X.control.ps)%*%(as.vector(ols.treat$coefficients)-as.vector(ol
#
#     # Compute ATEs
#     ate.ri.ll       = mean(c(tvec.ri.treat.ll,tvec.ri.control.ll))
#     ate.ri.ps       = mean(c(tvec.ri.treat.ps,tvec.ri.control.ps))
#
#     # Compute ATT
#     att.ri          = mean(tvec.ri.treat.ll)
#
#
#
#   }
#



#=============================================================#
# ==== IPW and Doubly Robust using the "CausalGAM" package ====
```

```r
#==============================================================#

  #========================#
  # ==== run functions ====
  #========================#


    # Covariates A, Lalonde control
      # make formula
      pscore_form <- as.formula(paste0("treat~", paste(z.a, collapse = " + ")))
      out_form    <- as.formula(paste0("re78~", paste(z.a, collapse = " + ")))
    ATE.ll.A <- estimate.ATE(pscore.formula = pscore_form,
                             pscore.family = binomial,
                             outcome.formula.t = out_form,
                             outcome.formula.c =out_form,
                             outcome.family = gaussian,
                             treatment.var = "treat",
                             data=as.data.frame(lal_c1),
                             divby0.action="t",
                             divby0.tol=0.001,
                             var.gam.plot=FALSE,
                             nboot=0)

    # Covariates B, Lalonde control
    pscore_form <- as.formula(paste0("treat~", paste(z.b, collapse = " + ")))
    out_form    <- as.formula(paste0("re78~", paste(z.b, collapse = " + ")))
    ATE.ll.B <- estimate.ATE(pscore.formula = pscore_form,
                             pscore.family = binomial,
                             outcome.formula.t = out_form,
                             outcome.formula.c =out_form,
                             outcome.family = gaussian,
                             treatment.var = "treat",
                             data=as.data.frame(lal_c1),
                             divby0.action="t",
                             divby0.tol=0.001,
                             var.gam.plot=FALSE,
                             nboot=0)

    # Covariates C, Lalonde control
    pscore_form <- as.formula(paste0("treat~", paste(z.c, collapse = " + ")))
    out_form    <- as.formula(paste0("re78~", paste(z.c, collapse = " + ")))
    ATE.ll.C <- estimate.ATE(pscore.formula = pscore_form,
                             pscore.family = binomial,
                             outcome.formula.t = out_form,
                             outcome.formula.c =out_form,
                             outcome.family = gaussian,
                             treatment.var = "treat",
                             data=as.data.frame(lal_c1),
                             divby0.action="t",
                             divby0.tol=0.001,
                             var.gam.plot=FALSE,
                             nboot=0)
    # # This doesnt run
```

```r
  # # Covariates A, PSID control
  # # make formula
  # pscore_form <- as.formula(paste0("treat~", paste(z.a, collapse = " + ")))
  # out_form    <- as.formula(paste0("re78~", paste(z.a, collapse = " + ")))
  # ATE.ps.A <- estimate.ATE(pscore.formula = pscore_form,
  #                          pscore.family = binomial,
  #                          outcome.formula.t = out_form,
  #                          outcome.formula.c =out_form,
  #                          outcome.family = gaussian,
  #                          treatment.var = "treat",
  #                          data=as.data.frame(lal_c2),
  #                          divby0.action="t",
  #                          divby0.tol=0.01,
  #                          var.gam.plot=FALSE,
  #                          nboot=0)

  # Covariates B, PSID control
  pscore_form <- as.formula(paste0("treat~", paste(z.b, collapse = " + ")))
  out_form    <- as.formula(paste0("re78~", paste(z.b, collapse = " + ")))
  ATE.ps.B <- estimate.ATE(pscore.formula = pscore_form,
                          pscore.family = binomial,
                          outcome.formula.t = out_form,
                          outcome.formula.c =out_form,
                          outcome.family = gaussian,
                          treatment.var = "treat",
                          data=as.data.frame(lal_c2),
                          divby0.action="t",
                          divby0.tol=0.001,
                          var.gam.plot=FALSE,
                          nboot=0)

  # Covariates C, PSID control
  pscore_form <- as.formula(paste0("treat~", paste(z.c, collapse = " + ")))
  out_form    <- as.formula(paste0("re78~", paste(z.c, collapse = " + ")))
  ATE.ps.C <- estimate.ATE(pscore.formula = pscore_form,
                          pscore.family = binomial,
                          outcome.formula.t = out_form,
                          outcome.formula.c =out_form,
                          outcome.family = gaussian,
                          treatment.var = "treat",
                          data=as.data.frame(lal_c2),
                          divby0.action="t",
                          divby0.tol=0.001,
                          var.gam.plot=FALSE,
                          nboot=0)



#=======================#
# ==== sort output ====
#=======================#

  #NOTE this was a dumb way to do this but it is what it is
```

```r
  #==========================#
  # ==== reg imputation ====
  #==========================#


    # Reg imputation results
    reg_imp_fun <- function(reg_output, in_spec, in_cont){
    tbl  =   data.table(estimate = reg_output$ATE.reg.hat,
                        std.error = reg_output$ATE.reg.asymp.SE,
                        CI_L = reg_output$ATE.reg.hat-1.96*reg_output$ATE.reg.asymp.SE,
                        CI_U = reg_output$ATE.reg.hat+1.96*reg_output$ATE.reg.asymp.SE)
    tbl <- round(tbl,2)
    tbl[, CI := paste0("(", CI_L, ", ", CI_U, ")")]
    tbl[, CI_L := NULL]
    tbl[, CI_U := NULL]
    setnames(tbl, colnames(tbl), paste0(colnames(tbl),"_", in_cont )  )
    tbl[, specification := in_spec]

    return(tbl[])
    }

    # do it with esp data
    out_RI <- list()
    out_RI[["a"]] <- reg_imp_fun(ATE.ll.A, "a", "exp")
    out_RI[["b"]] <- reg_imp_fun(ATE.ll.B, "b", "exp")
    out_RI[["c"]] <- reg_imp_fun(ATE.ll.C, "c", "exp")

    out_RI <-rbindlist(out_RI)

    # now do it with PSID
    out_RI2 <- list()
    out_RI2[["b"]] <- reg_imp_fun(ATE.ps.B, "b", "PSID")
    out_RI2[["c"]] <- reg_imp_fun(ATE.ps.C, "c", "PSID")

    out_RI2 <-rbindlist(out_RI2)

    # merge them
    out_RI <- merge(out_RI, out_RI2, by = "specification", all = TRUE)

#==============#
# ==== IPW ====
#==============#

    # ipw results
    ipw_fun <- function(reg_output, in_spec, in_cont){
      tbl  =   data.table(estimate = reg_output$ATE.IPW.hat,
                          std.error = reg_output$ATE.IPW.asymp.SE,
                          CI_L = reg_output$ATE.IPW.hat-1.96*reg_output$ATE.IPW.asymp.SE,
                          CI_U = reg_output$ATE.IPW.hat+1.96*reg_output$ATE.IPW.asymp.SE)
      tbl <- round(tbl,2)
      tbl[, CI := paste0("(", CI_L, ", ", CI_U, ")")]
      tbl[, CI_L := NULL]
      tbl[, CI_U := NULL]
```

```r
    setnames(tbl, colnames(tbl), paste0(colnames(tbl),"_", in_cont )  )
    tbl[, specification := in_spec]

    return(tbl[])
  }

  # do it with esp data
  out_IPW <- list()
  out_IPW[["a"]] <- ipw_fun(ATE.ll.A, "a", "exp")
  out_IPW[["b"]] <- ipw_fun(ATE.ll.B, "b", "exp")
  out_IPW[["c"]] <- ipw_fun(ATE.ll.C, "c", "exp")

  out_IPW <-rbindlist(out_IPW)

  # now do it with PSID
  out_IPW2 <- list()
  out_IPW2[["b"]] <- ipw_fun(ATE.ps.B, "b", "PSID")
  out_IPW2[["c"]] <- ipw_fun(ATE.ps.C, "c", "PSID")

  out_IPW2 <-rbindlist(out_IPW2)

  # merge them
  out_IPW <- merge(out_IPW, out_IPW2, by = "specification", all = TRUE)


#=================================#
# ==== Doubly robust results ====
#=================================#


  # DR results
  DR_fun <- function(reg_output, in_spec, in_cont){
    tbl  =   data.table(estimate = reg_output$ATE.AIPW.hat,
                        std.error = reg_output$ATE.IPW.asymp.SE,
                        CI_L = reg_output$ATE.AIPW.hat-1.96*reg_output$ATE.AIPW.asymp.SE,
                        CI_U = reg_output$ATE.AIPW.hat+1.96*reg_output$ATE.AIPW.asymp.SE)
    tbl <- round(tbl,2)
    tbl[, CI := paste0("(", CI_L, ", ", CI_U, ")")]
    tbl[, CI_L := NULL]
    tbl[, CI_U := NULL]
    setnames(tbl, colnames(tbl), paste0(colnames(tbl),"_", in_cont )  )
    tbl[, specification := in_spec]

    return(tbl[])
  }



  # do it with esp data
  out_dr <- list()
  out_dr[["a"]] <- DR_fun(ATE.ll.A, "a", "exp")
  out_dr[["b"]] <- DR_fun(ATE.ll.B, "b", "exp")
  out_dr[["c"]] <- DR_fun(ATE.ll.C, "c", "exp")
```

```
      out_dr <-rbindlist(out_dr)

      # now do it with PSID
      out_dr2 <- list()
      out_dr2[["b"]] <- DR_fun(ATE.ps.B, "b", "PSID")
      out_dr2[["c"]] <- DR_fun(ATE.ps.C, "c", "PSID")

      out_dr2 <-rbindlist(out_dr2)

      # merge them
      out_dr <- merge(out_dr, out_dr2, by = "specification", all = TRUE)




#============================#
# ==== CONSTRUCT TABLE 1 ====
#============================#

# fill in statistic and stack data
out_dmeans[, statistic := "Mean Diff"]
ols_out[, statistic := "OLS"]
out_RI[, statistic := "Reg. Impute"]
out_IPW[, statistic := "IPW"]
out_dr[, statistic := "D. Robust"]

# stack them all
out_table <-rbind(out_dmeans, ols_out, out_RI, out_IPW, out_dr, fill = TRUE )

# set the column order
setcolorder(out_table, c("statistic", "specification", setdiff(colnames(out_table), c("statistic", "spe

# save it
write.csv(out_table, "C://Users/Nmath_000/Documents/Code/courses/econ 675/PS_4_tex/q2_results_R.csv", r


#============================================================#
# ==== now load in CSV and make the friggin latex table ====
#============================================================#

# load ATE
ATE_table <- fread("C://Users/Nmath_000/Documents/Code/courses/econ 675/PS_4_tex/Table1_ATE_resultq.csv




print(xtable(ATE_table, type = "latex"),
      file = "C://Users/Nmath_000/Documents/Code/courses/econ 675/PS_4_tex/q2table.tex",
      include.rownames = FALSE,
      floating = FALSE)

# ATT
att_table <- fread("C://Users/Nmath_000/Documents/Code/courses/econ 675/PS_4_tex/Table1_ATT_resultq.csv
```

```r
print(xtable(ATE_table, type = "latex"),
      file = "C://Users/Nmath_000/Documents/Code/courses/econ 675/PS_4_tex/q2table_att.tex",
      include.rownames = FALSE,
      floating = FALSE)

# can't get this shit to work
# latex(round(ATE_table, 2),
#       file=paste0("C://Users/Nmath_000/Documents/Code/courses/econ 675/PS_4_tex/q2table.tex"),
#       append=FALSE,
#       table.env=FALSE
#       ,center="none",
#       title="",
#       n.cgroup=c(4, 4),
#       cgroup=c("Experimental Data", "PSID Control"),
#       colheads=c("$\\hat{\\tau}$", "s.e.", "C.I.", "", "$\\hat{\\tau}$", "s.e.", "C.I.", ""),
#       n.rgroup=c(1, rep(3, 6)),
#       rgroup=c("Mean Diff.", "OLS", "Reg. Impute", "IPW", "D. Robust", "N1 Match", "p Match"),
#       rowname=c("", rep(c("a", "b", "c"), 8)))



#=====================#
# ==== question 3 ====
#=====================#

# clear workspace

rm(list = ls(pos = ".GlobalEnv"), pos = ".GlobalEnv")
cat("\f")

# set attributes for plot to default ea theme
plot_attributes <- theme( plot.background = element_rect(fill = "lightgrey"),
                          panel.grid.major.x = element_line(color = "gray90"),
                          panel.grid.minor  = element_blank(),
                          panel.background = element_rect(fill = "white", colour = "black") ,
                          panel.grid.major.y = element_line(color = "gray90"),
                          text = element_text(size= 30),
                          plot.title = element_text(vjust=0, hjust = 0.5, colour = "#0B6357",face = "bo

######################################################################
# Generate random data and simulate
######################################################################

N     = 50
M     = 1000
SIGMA = matrix(c(1,0.85,0.85,1),2,2)

set.seed(1234)

# Generate covariates
W     = replicate(M,rmvnorm(N, mean = c(0,0), sigma = SIGMA, method="chol"))

# Generate errors
```

```
E     = replicate(M,rnorm(50))

# Generate outcomes
Y     = sapply(1:M,function(i) rep(1,N)+W[,,i]%*%c(0.5,1)+E[,i])

# Get beta.hats
beta.hats = sapply(1:M,function(i) lm(Y[,i]~W[,,i])$coefficients[2])

# Get t-stats for gamma.hats
t.stats   = sapply(1:M,function(i) summary(lm(Y[,i]~W[,,i]))[["coefficients"]][, "t value"][3])

# Get beta.tildes
beta.tildes = sapply(1:M,function(i) lm(Y[,i]~W[,1,i])$coefficients[2])

# Construct betas if the model selection is used
beta.sel    = ifelse(t.stats>=1.96,beta.hats,beta.tildes)

#==========================================================#
# ==== [1] Summary Statistics for the different betas ====
#==========================================================#


# Summary statistics
beta.sum = data.table(rbind(summary(beta.hats),summary(beta.tildes),summary(beta.sel)))

# estimates
beta.sum[, estimator := c("i", "ii", "iii")]

# put in order
setcolorder(beta.sum, c("estimator", setdiff(colnames(beta.sum), "estimator")))

# Make kernenl desity plot
plot.dat = data.frame(beta = c(beta.hats,beta.tildes,beta.sel),Estimator=rep(c("hat", "tilde","sel"), ea

densplot = ggplot(plot.dat,aes(x=beta,fill=Estimator))+
  geom_density(alpha=0.5, kernel="e",bw="ucv")+
  ggtitle("Kernel Density Plots")+
  xlab("Point Estimator")+
  ylab("Density")+
  plot_attributes +
  scale_fill_discrete(
    name="Estimator",
    breaks=c("hat", "tilde", "sel"),
    labels=c("(i)", "(ii)", "(iii)"))+
  theme(legend.justification = c(0.05, 0.98), legend.position = c(0.05, 0.98))

# save summary stats and plot
print(xtable(beta.sum, type = "latex"),
      file = "C://Users/Nmath_000/Documents/Code/courses/econ 675/PS_4_tex/q3_sum_stats.tex",
      include.rownames = FALSE,
      floating = FALSE)

png(paste0("c:/Users/Nmath_000/Documents/Code/courses/econ 675/PS_4_tex/q4_den.png"), height = 800, wid
```

```
print(densplot)
dev.off()


########################################################################
# [2] Coverage rates
########################################################################

# Compute coverage rate for beta.hat
beta.hats.se       = sapply(1:M,function(i) summary(lm(Y[,i]~W[,,i]))[["coefficients"]][, "Std. Error"]
beta.hats.CIs      = cbind(beta.hats-1.96*beta.hats.se,beta.hats+1.96*beta.hats.se)
beta.hats.covered  = ifelse(0.5>=beta.hats.CIs[,1]&0.5<=beta.hats.CIs[,2],1,0)
beta.hat.cr        = mean(beta.hats.covered)

# Compute coverage rate for beta.tilde
beta.tildes.se       = sapply(1:M,function(i) summary(lm(Y[,i]~W[,1,i]))[["coefficients"]][, "Std. Erro
beta.tildes.CIs      = cbind(beta.tildes-1.96*beta.tildes.se,beta.tildes+1.96*beta.tildes.se)
beta.tildes.covered  = ifelse(0.5>=beta.tildes.CIs[,1]&0.5<=beta.tildes.CIs[,2],1,0)
beta.tilde.cr        = mean(beta.tildes.covered)

# Compute coverage rate for beta.sel
beta.sel.CI.lower    = ifelse(beta.hats==beta.sel,beta.hats-1.96*beta.hats.se,beta.tildes-1.96*beta.til
beta.sel.CI.upper    = ifelse(beta.hats==beta.sel,beta.hats+1.96*beta.hats.se,beta.tildes+1.96*beta.til
beta.sel.CIs         = cbind(beta.sel.CI.lower,beta.sel.CI.upper)
beta.sel.covered     = ifelse(0.5>=beta.sel.CIs[,1]&0.5<=beta.sel.CIs[,2],1,0)
beta.sel.cr          = mean(beta.sel.covered)

# Put results together
cr.results           = rbind(beta.hat.cr,beta.tilde.cr,beta.sel.cr)
rownames(cr.results) = c("beta.hat.cr","beta.tilde.cr","beta.sel.cr")
colnames(cr.results) = c("Coverage Rate")

# save this shiz
print(xtable(cr.results, type = "latex"),
      file = "C://Users/Nmath_000/Documents/Code/courses/econ 675/PS_4_tex/q3_cov_rate.tex",
      include.rownames = FALSE,
      floating = FALSE)
```

4.2 STATA Code

```
*************************************************************************
* Preliminaries
*************************************************************************
clear all
set more off


*************************************************************************
* Import data, create additional covariates
*************************************************************************

* Import LaLonde data
import delimited using "C:\\Users\Nmath_000\Documents\MI_school\Second Year\675 Applied
Econometrics\hw\hw4\LaLonde_all.csv"


* set directory
cd "C:\Users\Nmath_000\Documents\Code\courses\econ 675\PS_4_tex\"
* Generate additional covariates
gen log_re74 = log(re74+1)
gen log_re75 = log(re75+1)
gen age_sq   = age^2
gen age_cu   = age^3
gen educ_sq  = educ^2
gen black_u74 = black*u74
gen educ_log_re74 = educ*log_re74
gen treat2    = treat if treat==1|treat==2
replace treat2=0 if treat2==2

*************************************************************************
* [1] Difference in means
*************************************************************************

* Lalonde control
reg re78 treat if treat==1|treat==0 , hc2

* PSID control
reg re78 treat if treat==1|treat==2 , hc2

*************************************************************************
* [2] OLS
*************************************************************************

* Covariates A, Lalonde control
reg re78 treat age educ black hisp married nodegr log_re74 log_re75 if treat==1|treat==0 ,
hc2

* Covariates B, Lalonde control
reg re78 treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75
if treat==1|treat==0 , hc2

* Covariates C, Lalonde control
reg re78 treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75
age_cu black_u74 educ_log_re74 if treat==1|treat==0 , hc2

* Covariates A, PSID
reg re78 treat age educ black hisp married nodegr log_re74 log_re75 if treat==1|treat==2 ,
hc2

* Covariates B, PSID
reg re78 treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75
if treat==1|treat==2 , hc2

* Covariates C, PSID
reg re78 treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75
age_cu black_u74 educ_log_re74 if treat==1|treat==2 , hc2
```

```
********************************************************************************
* [3] Regression Imputation
********************************************************************************

* Covariates A, Lalonde control
teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75) (treat) if treat==1|treat==0 , ate
teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75) (treat) if treat==1|treat==0 , atet

* Covariates B, Lalonde control
teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75) (treat) if treat==1|treat==0 , ate
teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75) (treat) if treat==1|treat==0 , atet

* Covariates C, Lalonde control
teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74) (treat) if treat==1|treat==0 , ate
teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74) (treat) if treat==1|treat==0 , atet


* Covariates A, PSID control
eststo ri1: teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75) (treat2) if treat2==1|treat2==0 , ate
eststo ri2: teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75) (treat2) if treat2==1|treat2==0 , atet

* Covariates B, PSID control
teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75) (treat2) if treat2==1|treat2==0 , ate
eststo ri3: teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75) (treat2) if treat2==1|treat2==0 , atet

* Covariates C, PSID control
eststo ri4: teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74) (treat2) if treat2==1|treat2==0 , ate
eststo ri5: teffects ra (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74) (treat2) if treat2==1|treat2==0 , atet

esttab ri1 using Q2_atematch.csv, se nostar keep(r1vs0.treat2) wide noparentheses nonumber noobs plain nomtitles replace
esttab ri2 ri3 ri4 using Q2_att.csv, se nostar keep(r1vs0.treat2) wide noparentheses nonumber noobs plain nomtitles replace

********************************************************************************
* [4] IPW
********************************************************************************

* Covariates A, Lalonde control
teffects ipw (re78) (treat age educ black hisp married nodegr log_re74 log_re75, logit) if treat==1|treat==0 , ate
teffects ipw (re78) (treat age educ black hisp married nodegr log_re74 log_re75, logit) if treat==1|treat==0 , atet

* Covariates B, Lalonde control
teffects ipw (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat==1|treat==0 , ate
teffects ipw (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat==1|treat==0 , atet

* Covariates C, Lalonde control
teffects ipw (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat==1|treat==0 , ate
teffects ipw (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat==1|treat==0 , atet

* Covariates A, PSID control [doesn't converge, so set maxiter = 50!!!]
eststo i1: teffects ipw (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75,
```

```
 logit) if treat2==1|treat2==0 , ate iterate(25)
eststo i2: teffects ipw (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75, logit) if treat2==1|treat2==0 , atet iterate(25)

* Covariates B, PSID control [first need to drop obs with very low prop scores]
teffects ipw (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat2==1|treat2==0 , ate osample(viol)
teffects ipw (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat2==1|treat2==0 & viol==0 , ate iter(25)
eststo i3: teffects ipw (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat2==1|treat2==0 & viol==0 , atet iter(25)

* Covariates C, PSID control [need to drop people]
teffects ipw (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat2==1|treat2==0 , ate osample(viol1)
teffects ipw (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat2==1|treat2==0 & viol1==0, ate iter(25)
eststo i4: teffects ipw (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat2==1|treat2==0 , atet iter(25)

esttab i1 using Q2_atematch.csv, se nostar keep(r1vs0.treat2) wide noparentheses nonumber noobs plain nomtitles append
esttab i2 i3 i4 using Q2_att.csv, se nostar keep(r1vs0.treat2) wide noparentheses nonumber noobs plain nomtitles append

*************************************************************************************
* [5] Doubly Robust
*************************************************************************************

* Covariates A, Lalonde control
teffects ipwra (re78) (treat age educ black hisp married nodegr log_re74 log_re75, logit) if treat==1|treat==0 , ate
teffects ipwra (re78) (treat age educ black hisp married nodegr log_re74 log_re75, logit) if treat==1|treat==0 , atet

* Covariates B, Lalonde control
teffects ipwra (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat==1|treat==0 , ate
teffects ipwra (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat==1|treat==0 , atet

* Covariates C, Lalonde control
teffects ipwra (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat==1|treat==0 , ate
teffects ipwra (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat==1|treat==0 , atet

* Covariates A, PSID control
eststo d1: teffects ipwra (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75, logit) if treat2==1|treat2==0 , ate iter(25)
eststo d2: teffects ipwra (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75, logit) if treat2==1|treat2==0 , atet iter(25)

* Covariates B, PSID control
teffects ipwra (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat2==1|treat2==0 , ate iter(25)
eststo d3: teffects ipwra (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat2==1|treat2==0 , atet iter(25)

* Covariates C, PSID control
teffects ipwra (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat2==1|treat2==0 , ate
eststo d4: teffects ipwra (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat2==1|treat2==0 , atet iter(25)

esttab d1 using Q2_atematch.csv, se nostar keep(r1vs0.treat2) wide noparentheses nonumber
```

```
noobs plain nomtitles append
esttab d2 d3 d4 using Q2_att.csv, se nostar keep(r1vs0.treat2) wide noparentheses nonumber noobs plain nomtitles append

*******************************************************************************
* [6] Nearest Neighbour Matching
*******************************************************************************

* Covariates A, Lalonde control
eststo n1: teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75) (treat) if treat==1|treat==0 , ate nneighbor(1) metric(maha)
eststo n2: teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75) (treat) if treat==1|treat==0 , atet nneighbor(1) metric(maha)

* Covariates B, Lalonde control
eststo n3: teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75) (treat) if treat==1|treat==0 , ate nneighbor(1) metric(maha)
eststo n4: teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75) (treat) if treat==1|treat==0 , atet nneighbor(1) metric(maha)

* Covariates C, Lalonde control
eststo n5: teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74) (treat) if treat==1|treat==0 , ate nneighbor(1) metric(maha)
eststo n6: teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74) (treat) if treat==1|treat==0 , atet nneighbor(1) metric(maha)

* Covariates A, PSID control
eststo n7: teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75) (treat2) if treat2==1|treat2==0 , ate nneighbor(1) metric(maha)
eststo n8:teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75) (treat2) if treat2==1|treat2==0 , atet nneighbor(1) metric(maha)

* Covariates B, PSID control
eststo n9:teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75) (treat2) if treat2==1|treat2==0 , ate nneighbor(1) metric(maha)
eststo n10:teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75) (treat2) if treat2==1|treat2==0 , atet nneighbor(1) metric(maha)

* Covariates C, PSID control
eststo n11:teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74) (treat2) if treat2==1|treat2==0 , ate nneighbor(1) metric(maha)
eststo n12:teffects nnmatch (re78 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74) (treat2) if treat2==1|treat2==0 , atet nneighbor(1) metric(maha)

esttab n7 using Q2_atematch.csv, se nostar keep(r1vs0.treat2) wide noparentheses nonumber noobs plain nomtitles append
esttab n8 n10 n12 using Q2_att.csv, se nostar keep(r1vs0.treat2) wide noparentheses nonumber noobs plain nomtitles append

*******************************************************************************
* [7] PS matching
*******************************************************************************

* Covariates A, Lalonde control
eststo p1: teffects psmatch (re78) (treat age educ black hisp married nodegr log_re74 log_re75, logit) if treat==1|treat==0 , ate
eststo p2: teffects psmatch (re78) (treat age educ black hisp married nodegr log_re74 log_re75, logit) if treat==1|treat==0 , atet

* Covariates B, Lalonde control
eststo p3: teffects psmatch (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat==1|treat==0 , ate
eststo p4: teffects psmatch (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat==1|treat==0 , atet

* Covariates C, Lalonde control
```

```
eststo p5: teffects psmatch (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat==1|treat==0 , ate
eststo p6: teffects psmatch (re78) (treat age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat==1|treat==0 , atet

* Covariates A, PSID control
eststo p7:teffects psmatch (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75, logit) if treat2==1|treat2==0 , ate
eststo p8:teffects psmatch (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75, logit) if treat2==1|treat2==0 , atet

* For the PSID samples below there are some prop scores too close to 1.
* First I need to run the treat2ment models, identify the respondents w/ problematic prop scores -- this will cause the code to break
* Then I drop the violators and estimate the treat2ment effects
teffects psmatch (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat2==1|treat2==0 , ate osample(viol2)
teffects psmatch (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat2==1|treat2==0, ate osample(viol3)


* Covariates B, PSID control
eststo p9:teffects psmatch (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat2==1|treat2==0 & viol2==0 , ate
eststo p10:teffects psmatch (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75, logit) if treat2==1|treat2==0 & viol2==0, atet

* Covariates C, PSID control
eststo p11: teffects psmatch (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat2==1|treat2==0 & viol3==0 , ate
eststo p12: teffects psmatch (re78) (treat2 age educ black hisp married nodegr log_re74 log_re75 age_sq educ_sq u74 u75 age_cu black_u74 educ_log_re74, logit) if treat2==1|treat2==0 & viol3==0 , atet

esttab p1 p3 p5 p7 p9 n11 using Q2_atematch.csv, se nostar keep(r1vs0.treat r1vs0.treat2) wide noparentheses nonumber noobs plain nomtitles append
esttab p8 p10 p12 using Q2_att.csv, se nostar keep(r1vs0.treat2) wide noparentheses nonumber noobs plain nomtitles append





*****************************************************************************
* Preliminaries
*****************************************************************************
clear all
set more off

* Set working directory
global dir "/Users/Anirudh/Desktop/GitHub"


set seed 22
set obs 50

*****************************************************************************
* [1] Summary stats and density plots
*****************************************************************************

* number of replications
local M = 1000
set matsize 11000

* empty matrices to store estimates and indicator of coverage
matrix est = J(`M',3,.)
```

```
matrix cov = J(`M',3,.)

* initial values we will replace during replication
gen x = rnormal(0,1)
gen z = .85*x + sqrt(1-.85)*rnormal(0,1)
gen eps = rnormal(0,1)
gen y = 1 + .5*x + z + eps

* loop for M replications
forvalues i = 1/`M'{
    qui replace x = rnormal(0,1)
    qui replace z = .85*x + sqrt(1-.85)*rnormal(0,1)
    qui replace eps = rnormal(0,1)
    qui replace y = 1 + .5*x + z + eps

    * long regression
    qui reg y x z, r

    * extract first estimate
    local beta_hat = _b["x"]
    matrix est[`i',1] = `beta_hat'

    * get SE and calculate coverage of true beta_0 = .5
    local se_hat = _se["x"]
    local lb_hat = `beta_hat' - 1.96 * `se_hat'
    local ub_hat = `beta_hat' + 1.96 * `se_hat'
    local cov_hat = (.5 >= `lb_hat') & (.5 <= `ub_hat')
    matrix cov[`i',1] = `cov_hat'

    * save gamma over se gamma
    local gamma_hat = _b["z"]
    local gamma_se  = _se["z"]
    local tstat = `gamma_hat'/`gamma_se'

    * short regression
    qui reg y x, r
    local beta_tilde = _b["x"]
    matrix est[`i',2] = `beta_tilde'

    * get SE and calculate coverage of true beta_0 = .5
    local se_tilde = _se["x"]
    local lb_tilde = `beta_tilde' - 1.96 * `se_tilde'
    local ub_tilde = `beta_tilde' + 1.96 * `se_tilde'
    local cov_tilde = (.5 >= `lb_tilde') & (.5 <= `ub_tilde')
    matrix cov[`i',2] = `cov_tilde'

    * third estimate
    local beta_check = cond(`tstat' >= 1.96, `beta_hat', `beta_tilde')
    matrix est[`i',3] = cond(`tstat' >= 1.96, `beta_hat', `beta_tilde')
    matrix cov[`i',3] = cond(`tstat' >= 1.96, `cov_hat', `cov_tilde')
}

* turn results into variables
svmat est
svmat cov

* drop old data
drop x
drop z
drop eps
drop y

* rename variables
rename est1 beta_hat
rename est2 beta_tilde
rename est3 beta_check
rename cov1 cov_hat
rename cov2 cov_tilde
rename cov3 cov_check
```

```
* write summary statistics to latex
outreg2 using q3.tex, replace sum(log) ///
    keep(beta_hat beta_tilde beta_check) ///
    eqkeep(min mean median max) ///
    dec(2)

* kernel densities
twoway kdensity beta_hat, k(epanechnikov) || ///
 kdensity beta_tilde, k(epanechnikov) || ///
 kdensity beta_check, k(epanechnikov) ///
 leg(lab(1 "beta_hat") lab(2 "beta_tilde") lab(3 "beta_check")) ///
 ytitle("Density") xtitle("")


********************************************************************************
* [2] Coverage rates
********************************************************************************

* calculate these here, report them in LaTeX
sum(cov_hat)
sum(cov_tilde)
sum(cov_check)

```