Assignment_Two Q2

Author - Eric Opoku

November 11, 2022

```
name<- Sys.info()
name[7]
   user
"erico"
```

```
library(rlang)
library(dplyr)
library(kableExtra)
library(ggdag)      # For plotting DAGs
library(dagitty)    # For working with DAG logic
library(modelsummary) # For making regression tables
library(AER) # this package has lots of applied metrics packages
library(foreign)
# Helpful for reading in data from Stata or other code languages
library(lubridate) # For figures
library(stargazer) # For tables
library(data.table)  #data manipulation and wrangling
library(lme4)
library(psych)
library(readxl) # Read in data
library(expss) #value labelling from spss style
library(readstata13)
library(marginaleffects) # To calculate marginal effects
library(knitr) # Alternative table package
set.seed(03262020) # random number generators; same numbers across machines
```

```
#https://stat.ethz.ch/pipermail/r-help/2017-August/448710.html


load("C:/Users/erico/Desktop/ACADEMICS/MY UofT COURSES/YEAR TWO/SEMESTER ONE/HAD5744H Ap
```

```
Loading objects:
  addvars

file.exists("C:/Users/erico/Desktop/ACADEMICS/MY UofT COURSES/YEAR TWO/SEMESTER ONE/HAD5
[1] TRUE


ls()
[1] "addvars" "name"

mydata<-addvars

#library(readr)
library(R.utils)
saveObject(addvars, "C:/Users/erico/Desktop/ACADEMICS/MY UofT COURSES/YEAR TWO/SEMESTER
mydata1 <- loadObject("C:/Users/erico/Desktop/ACADEMICS/MY UofT COURSES/YEAR TWO/SEMESTE


#name of stored object is addvars


mydata2 <- read.dta("C:/Users/erico/Desktop/ACADEMICS/MY UofT COURSES/YEAR TWO/SEMESTER


mydata2<- select (mydata2, RCT_ID_TV2, post_ie_charges_180_IP,
                  Ireadmit2_180_100,
                  Iaa_100, Ihispanic_100, Iwhite_100, Irace_other_100)

mydata1<- select (mydata1, RCT_ID_TV2, link2care_duration, Icontrol, Itreatment,
                  Imale)


mydata<- merge(mydata1, mydata2, by = c("RCT_ID_TV2"), all =T)


mydata <- mydata %>% mutate(treated=ifelse(link2care_duration>89,1,0))
mydata <- mydata %>% mutate(random=ifelse(Itreatment==1,1,0))
mydata <- mydata %>%mutate(rand_treated=ifelse(random==1&treated==1,1,0))
mydata <- mydata %>% mutate(rand_untreated=ifelse(random==1&treated==0,1,0))
mydata <- mydata %>% mutate(male=ifelse(Imale==1,1,0))
mydata <- mydata %>% mutate(african_american=ifelse(Iaa_100==100,1,0))
mydata <- mydata %>% mutate(hispanic=ifelse(Ihispanic_100==100,1,0))
mydata <- mydata %>%mutate(white=ifelse(Iwhite_100 ==100,1,0))
mydata <- mydata %>% mutate(other=ifelse(Irace_other_100 ==100,1,0))
```

```
mydata <- mydata %>% mutate(readmit=ifelse(Ireadmit2_180_100==100,1,0))

#correct name
#mydata <- select and merge

 #merge the Dataset 2b RCt and also the RCT additional variables to create the;
#Ireadmit2_180_100 and post_ie_charges_180_IP
```

#Question A

Voluntary participation in the program introduces selection bias (or endogeneity) into the RCT. In the DAG, gender affects both program participation and the rate of 180-day readmission. It also affects the rate of 180-day readmission through program participation.

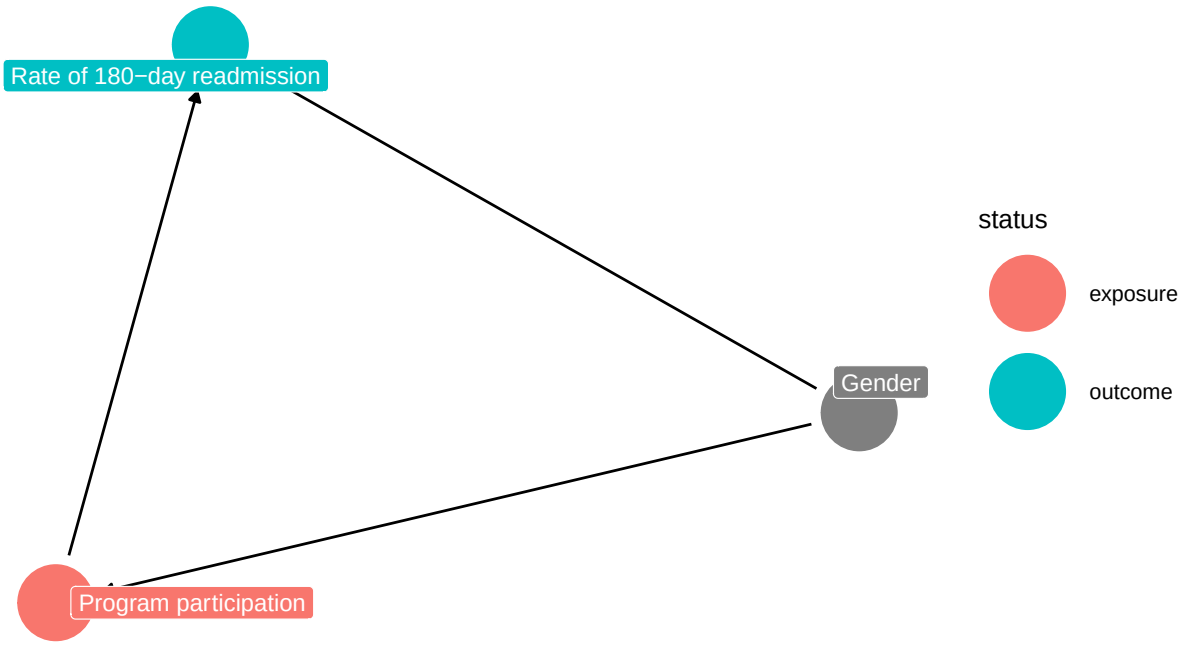
It is possible that there are unobserved characteristics of individuals that affect both individual program take up and the rate of hospital readmission and it may account for any association between program participation and the 180-day hospital readmission rates. For example, unhealthy individuals who have high likelihood of re-admissions might self-select themselves into the program and hence, its features might have little causal impact on their actions or behaviors and outcomes.

In addition, super-utilizers of health care or super-spenders who do not participate in the program might be fundamentally different on unobserved characteristics from those who actually participated. This process of program participation must be accounted for, if not, the observed association between program participation and 180-day readmission rates likely reflect the effect of unmeasured individual or family characteristics that motivated program participation.

It is worth mentioning here that since participation is voluntary, we may not be able to observe the potential outcome of individuals who do not participate in the program. Hence there will be sample selection.

In conclusion, voluntary participation biases the estimates and makes them inconsistent. The 'selection effects' must be separated from the 'program effects'.

```
the_dag <- dagify(  # Create super basic DAG
  readmit ~ male,
  rand_treated ~ male,
  readmit ~ rand_treated,
  exposure = "rand_treated",
  outcome = "readmit",
  labels = c(readmit = "Rate of 180-day readmission",
             rand_treated = "Program participation",
             male="Gender"))

set.seed(26445111)  #prevents DAG shape from changing
ggdag_status(the_dag, use_labels= "label",
             text = FALSE) + theme_dag()
```



#Question B

Using the potential outcome framework, let Z (i.e., the instrumental variable) represent whether an individual i was initially randomized to take up the program (Z=1) or not (Z=0). Let D denote the treatment- whether an individual i actually participated or took up the program (D=1) or not (D=0). D is a potentially endogenous variable in the model. Let Y denote the outcome- the rate of 180-day hospital readmission of individual i. Let X denote control variables or covariates.

Let Y_{i}(D_{i}=1) denote the potential outcome that would be observed for individual i if he/she had actually participated in the program. However, we would not be able to observe Y_{i}(D_{i}=1) for individual i if he/she did not actually participate in the program. Thus, Y_{i}(D_{i}=1) would denote unobserved, counterfactual outcome if he/she had not participated (control group). The treatment effect for individual i is the difference between his/her potential outcomes under treatment and control: Y_{i}(D_{i}=1)-Y_{i}(D_{i}=0). Meanwhile, only one of these potential outcomes is observed (Holland, 1986; Felton and Stewart, 2022).

Let D_{i}(Z_{i})=1) denote the potential treatment (i.e., actual program take up or not) that would be observed if individual i had been randomly assigned to the program rather than to the control group. Y_{i}(D_{i}=1, Z_{i})=1) represents the potential outcome that would be observed for individual i had he/she had been randomized into the program and had also actually taken up the program.

The identification of the randomization as instrumental variable is analyzed based on six assumptions: Relevance, Exclusion restriction, Exogeneity, unconfoundness, monotonicity, stableunit-treatment-value assumption (SUTVA) and positivity.

Relevance depicts that being randomized into the program has a causal effect on whether individual i participated in the program or not and such impact is expected in this RCT. Thus: E[D_{i}(Z_{i})=1)-D_{i}(Z_{i})=0)] is not equal to zero.

Unconfounded (or exogenous) instrument means that the instrument (i.e., randomization) can share no unmeasured common causes with neither the outcome (i..e, rate of 180-day readmission) nor the treatment (i.e., program participation). This assumption is also satisfied in the model as the IV is randomized.

The exclusion restriction requires that the instrumental variable has no effect on the outcome except through the treatment. Here, the randomization instrumental variable affects the rate of 180-day hospital readmission (i.e., outcome) only through the program participation (i.e., treatment).

Monotonicity: Let's identify four plausible "compliance types". Assume people are randomized to the control and treatment arms or groups. Hence, randomization is unconfounded since participants are assigned randomly.

Neuer is a *never taker* if, regardless of the group he is randomly *assigned* to, he is never going to *participate* in the program itself. Whitaker is a *always taker* if, regardless of the group he is randomly *assigned* to, he is always going to *participate* in the program. Collins is a *complier* if he will *participate* in the program if he is assigned to the treatment arm but

he will not participate if he was randomized into the control group. Daphne is a *defier* if she would participate in the program even though she was initially randomized not to join the treatment group but would not participate even if she was initially randomized to join the treatment group. The monotonicity assumption, which is untestable, expects that there are no *defiers* in the sample (Felton and Stewart, 2022). The SUTVA assumption expects no *hidden* versions of the instrument or treatment and strikes out interference across units.

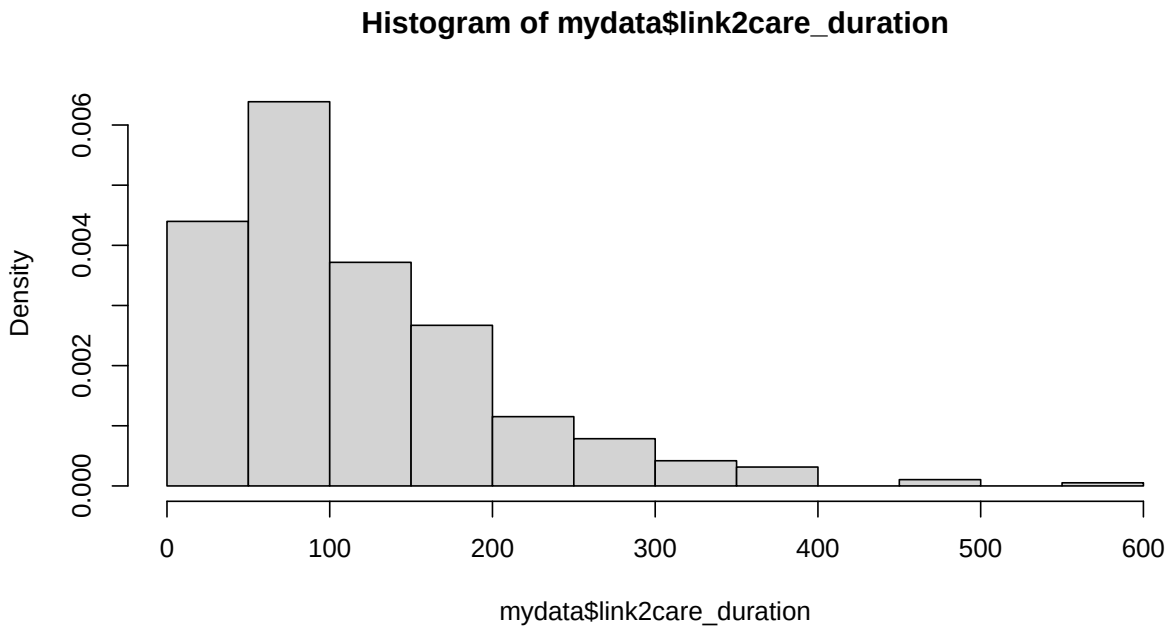
Positivity requires that, in every stratum of measured confounders (i.e., those necessary for unconfoundedness of the instrument to hold) at least some units receive the instrument (Aronow and Miller, 2019).

To the best of my knowledge, the randomization instrumental variable meets the testable assumptions.

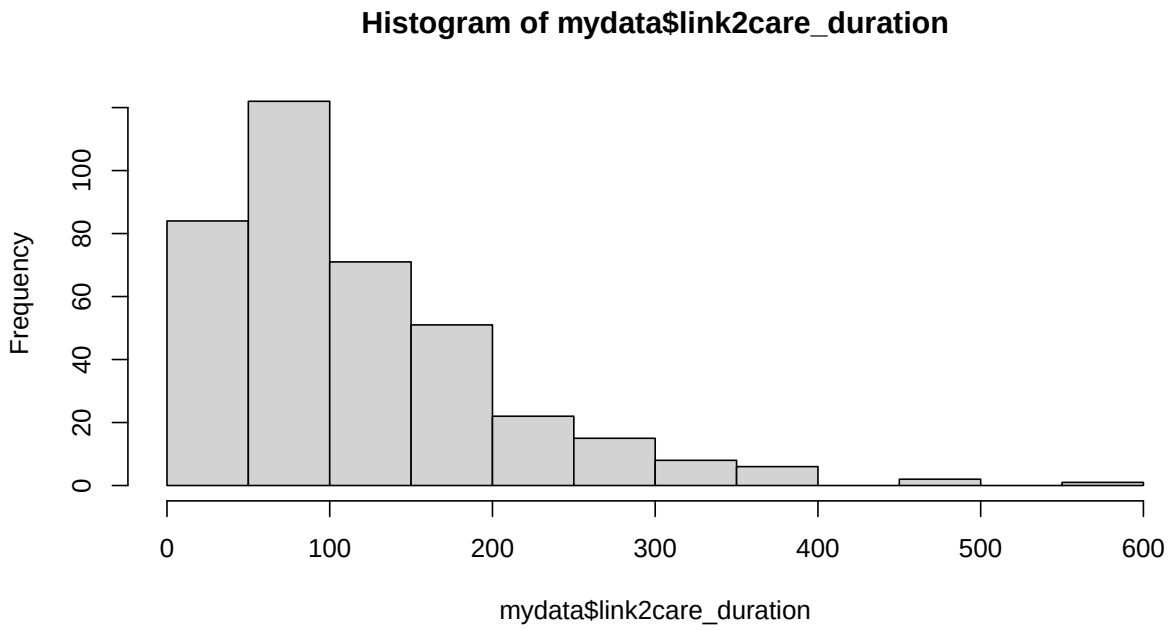
##Question C

The link2care_duration is left-skewed. There are lesser individuals who were randomized in the control group (n=389), a dummy variable, relative to those who were randomized into the treatment group (n=393). Those who were randomized and treated (n=194), a dummy variable, were less relative to their counterparts (n=577). Those who were randomized but not treated (n=188), a dummy variable, were lesser relative to their counterparts (n=583).

```
#mydata <- mydata %>% mutate(coverage = ifelse(q0411<4, 1, 0))
#summary(mydata$treated)
#table(mydata$treated)

#table(mydata$link2care_duration)
hist(mydata$link2care_duration, freq=F) #Density
```




```
hist(mydata$link2care_duration, freq=T) #Frequency
```




```
#summary of control groups
summary(mydata$Icontrol)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.0000  0.0000  0.4974  1.0000  1.0000
table(mydata$Icontrol, exclude = NA)

  0   1
393 389
describe(mydata$Icontrol)
   vars   n mean  sd median trimmed mad min max range skew kurtosis   se
X1    1 782  0.5 0.5      0     0.5   0   0   1     1 0.01       -2 0.02
str(mydata$Icontrol)
 int [1:782] 1 0 0 1 1 1 1 1 0 1 ...

#interact Itreatment with treated variable
summary(mydata$Itreatment)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.0000  1.0000  0.5026  1.0000  1.0000
#table(mydata$Itreatment)

#create randomization
#table(mydata$random)

#summary of randomized and treated
summary(mydata$rand_treated)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.0000  0.0000  0.0000  0.2516  1.0000  1.0000      11
table(mydata$rand_treated, exclude = NA)

  0   1
577 194
describe(mydata$rand_treated)
   vars   n mean   sd median trimmed mad min max range skew kurtosis   se
X1    1 771 0.25 0.43      0    0.19   0   0   1     1 1.14     -0.7 0.02
str(mydata$rand_treated)
 num [1:782] 0 0 1 0 0 0 0 0 1 0 ...


#summary of randomized but not treated
summary(mydata$rand_untreated)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.0000  0.0000  0.0000  0.2438  0.0000  1.0000      11
table(mydata$rand_untreated, exclude = NA)

  0   1
583 188
describe(mydata$rand_untreated)
   vars   n mean   sd median trimmed mad min max range skew kurtosis   se
X1    1 771 0.24 0.43      0    0.18   0   0   1     1 1.19    -0.58 0.02
str(mydata$rand_untreated)
 num [1:782] 0 1 0 0 0 0 0 0 0 0 ...

#confirmation with cross tabulation
#xtabs(~random+treated, data=mydata)

#desc_table

desc_table = data.frame(
  Measure = c("Control", "Randomized and Treated",
              "Randomized but not Treated"),
   M_1  = c(782, 771, 771),
  M_2  = c(389, 194, 188))


kable(
  desc_table,
  col.names = c("Measure","*Observation*", "*Number of Values*"),
  digits = 2,
  caption = "Summary of Variables"
```

Table 1: Summary of Variables

Measure	*Observation*	*Number of Values*
Control	782	389
Randomized and Treated	771	194
Randomized but not Treated	771	188

```
)
```

##QUESTION D

The coefficient (in the Logistic regression) means that there is a positive association between randomization and 'actual' take up or participation of the program. The F-statistic is greater than 104.7 hence the instrument is strong. This regression highlights the 'Relevance' assumption. Thus, the instrumental variable is expected to have significant effect on the treatment variable.

	First Stage-Logit
(Intercept)	−19.566***
	(2×10^{-8})
random	19.597***
	(0.108)
Num.Obs.	771
AIC	533.5
BIC	542.8
Log.Lik.	−264.735
F	33 005.295
RMSE	0.35
Std.Errors	HC3

* p < 0.1, ** p < 0.05, *** p < 0.01

```
table(mydata$treated)

  0   1
188 194

# First stage regression of IV: take up on randomization


m1 <- glm(rand_treated ~ random,data=mydata,
          family='binomial'(link='logit'))

msummary(list("First Stage-Logit"=m1),
        vcov=c("robust"),
         tab_header=c("First Stage"),
         stars=c('*' = .1, '**' = .05, '***' = .01))
```

##QUESTION E

There is not much difference between both the naive and the 2SLS models. Contrary to what was expected, the F statistics and *R-squared* for the naive model were higher than for the 2SLS model. Also, the estimates of the treatment variables is not significant for the 2SLS model but it is significant for the naive model (at 10 percent significance level). This shows that failure to account for potential endogeneity can lead to misleading results.

```
#first stage #logit #binary treatment
#random is the IV
#run endogenous variable on the IV and controls
#run that probit

m1r <- glm(rand_treated ~ random+african_american+hispanic+white+male,
           data=mydata,
          family='binomial'(link='probit'))

mydata$pred_x_X <- predict(m1r, mydata)

#Forbidden regresion #sticking predicted values into the 2nd stage
forbid <- lm(post_ie_charges_180_IP ~ pred_x_X+african_american+hispanic+white+
               male,data=mydata)


##The 2SLS #predicted values in place of instrument
m1_second <- lm(rand_treated ~ pred_x_X+african_american+hispanic+white+male,
                data=mydata)

mydata$pred_pred_x <- predict(m1_second, mydata)


#2 nd stage
# Regress outcome on pred_X
m3 <- lm(post_ie_charges_180_IP ~ pred_pred_x+african_american+hispanic+white+
           male,data=mydata)


#naive
m2 <- lm(post_ie_charges_180_IP ~ rand_treated+african_american+hispanic+white+
           male,data=mydata) #Linear model
msummary(list("Naive"=m2),
         vcov=c("robust"),
```

	Naive
(Intercept)	73 294.700
	(70 357.841)
rand_treated	32 476.350*
	(18 900.733)
african_american	28 962.300
	(71 138.197)
hispanic	41 090.264
	(71 583.065)
white	25 445.903
	(71 621.763)
male	5296.179
	(13 962.350)
Num.Obs.	771
R2	0.006
R2 Adj.	−0.0003
AIC	20 974.2
BIC	21 006.7
Log.Lik.	−10 480.092
F	0.837
RMSE	193 671.31
Std.Errors	HC3

* p < 0.1, ** p < 0.05, *** p < 0.01

```
         stars=c('*' = .1, '**' = .05, '***' = .01))
```

```
#Combination
msummary(list("Naive"=m2,"Forbidden"=forbid,"2SLS"=m3),
         vcov=c("robust","robust", "robust"),
         stars=c('*' = .1, '**' = .05, '***' = .01))
```

	Naive	Forbidden	2SLS
(Intercept)	73 294.700	89 867.209	87 665.330
	(70 357.841)	(74 146.775)	(74 913.654)
rand_treated	32 476.350*		
	(18 900.733)		
african_american	28 962.300	21 619.251	21 893.208
	(71 138.197)	(74 938.291)	(75 039.319)
hispanic	41 090.264	30 748.657	31 160.332
	(71 583.065)	(75 373.984)	(75 466.633)
white	25 445.903	13 593.149	13 958.747
	(71 621.763)	(75 437.049)	(75 545.228)
male	5296.179	7395.142	7372.826
	(13 962.350)	(13 870.551)	(13 853.843)
pred_x_X		314.619	
		(2284.640)	
pred_pred_x			3735.089
			(27 122.785)
Num.Obs.	771	782	782
R2	0.006	0.001	0.001
R2 Adj.	−0.0003	−0.005	−0.005
AIC	20 974.2	21 272.3	21 272.3
BIC	21 006.7	21 305.0	21 305.0
Log.Lik.	−10 480.092	−10 629.161	−10 629.161
F	0.837	0.256	0.256
RMSE	193 671.31	193 559.31	193 559.31
Std.Errors	HC3	HC3	HC3

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

#QUESTION F

An instrumental variable helps to recover the Local Average Treatment Effects (LATE). The LATE, also referred to as the complier average causal effect (CACE), denotes the treatment effect for the subset of the sample that participated in the program when they were initially randomized to do so (i.e., the compliers). This helps to fulfil the monotonicity assumption of eliminating *defiers*.

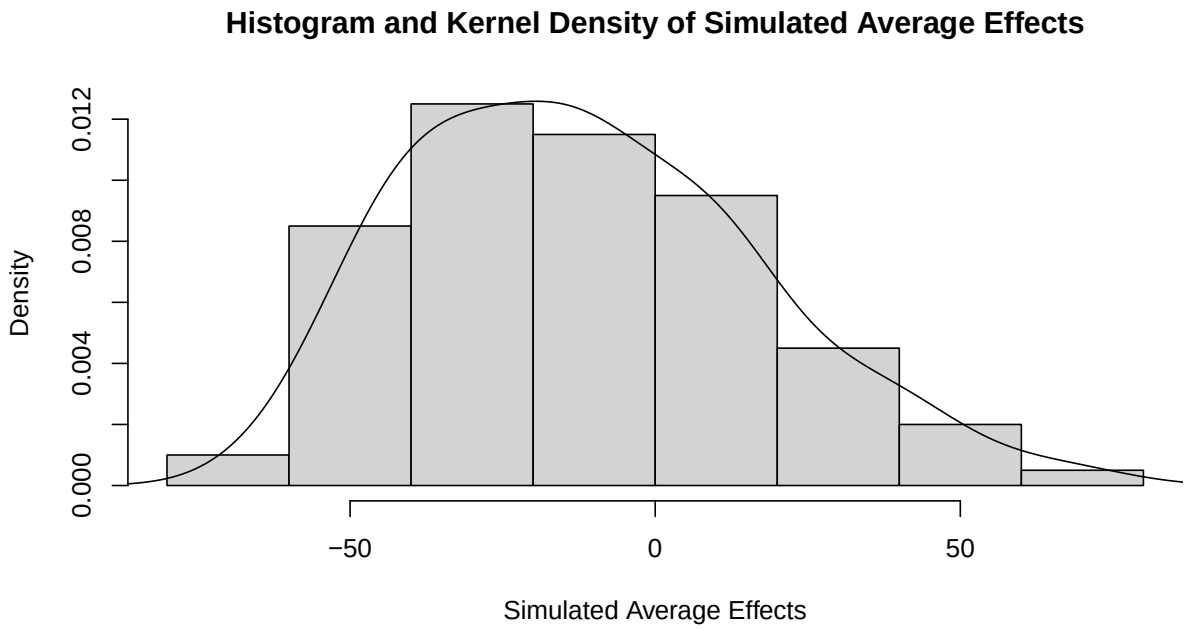
Meanwhile, the LATE is only policy or economically relevant if it has been proved in the literature that many individuals comply to such programs. There is not enough information to conclude on this so I would say that the LATE is not economically relevant.

#QUESTION G

The estimate for the treatment variable in the 'Naive-OLS' is significant (0.087) even though it is not significant in the 2SLS model. Their estimates show a positive association between program participation and the rate of 180-day hospital readmission. However, though insignificant, the estimates of the Bivariate regression shows a negative association between program participation and the rate of 180-day hospital readmission.

```
library(GJRM)
Loading required package: mgcv
Loading required package: nlme

Attaching package: 'nlme'
The following object is masked from 'package:lme4':

    lmList
The following object is masked from 'package:dplyr':

    collapse
This is mgcv 1.8-41. For overview type 'help("mgcv-package")'.

This is GJRM 0.2-6.
For overview type 'help("GJRM-package")'.
table(mydata$readmit)

  0   1
498 284

#male and race are the restricted variables
out2<-gjrm(list(rand_treated ~  random+african_american+hispanic+white,
               readmit ~ rand_treated+random+male),  #rand_treated
          data = mydata,
          margins = c("probit", "probit"),
          Model = "B")
conv.check(out2)

Largest absolute gradient value: 1.115763e-07
Observed information matrix is positive definite
Eigenvalue range: [1.348283e-07,908.8774]

Trust region iterations: 10
Estimated overall probability range: 1.997374e-13 0.9880989
Estimated overall density range: 1.804796e-11 1.088973
summary(out2)

COPULA:   Gaussian
MARGIN 1: Bernoulli
MARGIN 2: Bernoulli

EQUATION 1
Link function for mu.1: probit
Formula: rand_treated ~ random + african_american + hispanic + white
```

```
Parametric coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -6.2446  1925.7260  -0.003    0.997
random              6.7048  1925.7259   0.003    0.997
african_american   -0.3158     0.7537  -0.419    0.675
hispanic           -0.6574     0.7578  -0.868    0.386
white              -0.5430     0.7672  -0.708    0.479


EQUATION 2
Link function for mu.2: probit
Formula: readmit ~ rand_treated + random + male

Parametric coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.38223    0.07804  -4.898 9.68e-07 ***
rand_treated -0.24996    1.02893  -0.243    0.808
random        0.16429    0.57838   0.284    0.776
male          0.06390    0.09063   0.705    0.481
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


n = 771  theta = 0.353(-0.812,0.911)  tau = 0.229(-0.606,0.73)
total edf = 10

#AIC(out2); BIC(out2)
## Testing the hypothesis of absence of endogeneity post estimation...
#gt.bpm(out2)
## treatment effect, risk ratio and odds ratio with CIs and random effects
mb(mydata$rand_treated, mydata$Ireadmit2_180_100, Model = "B")

Worst-Case Manski's bounds for ATE (%):
Lower Bound: -40.2
Upper Bound: 59.8

Imbens&Manski's CI for ATE (%): [-40.3,59.9]

ATE assuming random assignment (%): 8.64
AT(out2, nm.end = "rand_treated", hd.plot = TRUE)
```




```
Treatment effect (%) with 95% interval:

-9.18 (-53.97,46.25)
RR(out2, nm.end = "rand_treated")

Risk ratio with 95% interval:

0.767 (0.109,3.556)
OR(out2, nm.end = "rand_treated")

Odds ratio with 95% interval:

0.6664 (0.0354,13.3539)
#AT(out2, nm.end = "rand_treated", type = "univariate")
#re.imp <- imputeCounter(out2, m = 10, "rand_treated")   #Random effects
#random effects
#re.imp$AT

#compare with below
# iii)

# ols #naive
m6_ols <- lm(readmit ~ rand_treated+african_american+hispanic+white+male,
             data=mydata)
```

```
# binary IV regression
m5_first <- glm(rand_treated ~ random+african_american+hispanic+white+male,
                data=mydata, family='binomial'(link='probit'))

#predict X with Z:
mydata$pred_x_naive <- predict(m5_first, mydata)

#replace IV with binnary IV prediction  #first stage
m5_first_naive <- lm(rand_treated ~ pred_x_naive+african_american+hispanic+white
                     +male,data=mydata)

#predict X with Z:
mydata$the_pred_x_naive <- predict(m5_first_naive, mydata)


# Regress outcome on pred_X
m6_second_naive <- lm(readmit ~ the_pred_x_naive+african_american+hispanic
                      +white+male,data=mydata)

msummary(list("Naive-OLS"=m6_ols,"Naive-2SLS"=m6_second_naive),
         vcov=c("robust","robust"),
         stars=c('*' = .1, '**' = .05, '***' = .01))
```

	Naive-OLS	Naive-2SLS
(Intercept)	0.206	0.249
	(0.273)	(0.290)
rand_treated	0.087**	
	(0.041)	
african_american	0.120	0.093
	(0.275)	(0.290)
hispanic	0.120	0.092
	(0.275)	(0.291)
white	0.181	0.146
	(0.277)	(0.293)
male	0.023	0.027
	(0.035)	(0.035)
the_pred_x_naive		0.002
		(0.069)
Num.Obs.	771	782
R2	0.009	0.003
R2 Adj.	0.003	−0.004
AIC	1070.8	1086.3
BIC	1103.4	1118.9
Log.Lik.	−528.413	−536.154
F	1.343	0.369
RMSE	0.48	0.48
Std.Errors	HC3	HC3

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

#QUESTION H

The results suggest that there is a negative association (insignificant) between take up of the program and the rate of 180-day readmission (for the Bivariate regression model). Hence, it can be deduced that the program is not effective in reducing the rate of 180-day readmission.

In addition, the 2SLS also showed that there is an insignificant association between program participation and superutilization spending. Hence, it can be inferred that the program is not effective in reducing superutilization spending.

```
packages <- knitr::write_bib(file = 'packages.bib')
packages
```

$AER @Manual{R-AER, title = {AER: Applied Econometrics with R}, author = {Christian Kleiber and Achim Zeileis}, year = {2022}, note = {R package version 1.2-10}, url = {https://CRAN.R-project.org/package=AER}, }

$base @Manual{R-base, title = {R: A Language and Environment for Statistical Computing}, author = {{R Core Team}}, organization = {R Foundation for Statistical Computing}, address = {Vienna, Austria}, year = {2022}, url = {https://www.R-project.org/}, }

$car @Manual{R-car, title = {car: Companion to Applied Regression}, author = {John Fox and Sanford Weisberg and Brad Price}, year = {2022}, note = {R package version 3.1-1}, url = {https://CRAN.R-project.org/package=car}, }

$carData @Manual{R-carData, title = {carData: Companion to Applied Regression Data Sets}, author = {John Fox and Sanford Weisberg and Brad Price}, year = {2022}, note = {R package version 3.0-5}, url = {https://CRAN.R-project.org/package=carData}, }

$dagitty @Manual{R-dagitty, title = {dagitty: Graphical Analysis of Structural Causal Models}, author = {Johannes Textor and Benito {van der Zander} and Ankur Ankan}, year = {2021}, note = {R package version 0.3-1}, url = {https://CRAN.R-project.org/package=dagitty}, }

$data.table @Manual{R-data.table, title = {data.table: Extension of `data.frame`}, author = {Matt Dowle and Arun Srinivasan}, year = {2022}, note = {R package version 1.14.4}, url = {https://CRAN.R-project.org/package=data.table}, }

$dplyr @Manual{R-dplyr, title = {dplyr: A Grammar of Data Manipulation}, author = {Hadley Wickham and Romain François and Lionel Henry and Kirill Müller}, year = {2022}, note = {R package version 1.0.10}, url = {https://CRAN.R-project.org/package=dplyr}, }

$expss @Manual{R-expss, title = {expss: Tables, Labels and Some Useful Functions from Spreadsheets and SPSS Statistics}, author = {Gregory Demin}, year = {2022}, note = {R package version 0.11.2}, url = {https://gdemin.github.io/expss/}, }

$foreign @Manual{R-foreign, title = {foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...}, author = {{R Core Team}}, year = {2022}, note = {R package version 0.8-83}, url = {https://svn.r-project.org/R-packages/trunk/foreign/}, }

$ggdag @Manual{R-ggdag, title = {ggdag: Analyze and Create Elegant Directed Acyclic Graphs}, author = {Malcolm Barrett}, year = {2022}, note = {R package version 0.2.7}, url = {https://github.com/malcolmbarrett/ggdag}, }

$GJRM @Manual{R-GJRM, title = {GJRM: Generalised Joint Regression Modelling}, author = {Giampiero Marra and Rosalba Radice}, year = {2022}, note = {R package version 0.2-6}, url = {https://www.ucl.ac.uk/statistics/people/giampieromarra}, }

$kableExtra @Manual{R-kableExtra, title = {kableExtra: Construct Complex Table with kable and Pipe Syntax}, author = {Hao Zhu}, year = {2021}, note = {R package version 1.3.4}, url = {https://CRAN.R-project.org/package=kableExtra}, }

$knitr @Manual{R-knitr, title = {knitr: A General-Purpose Package for Dynamic Report Generation in R}, author = {Yihui Xie}, year = {2022}, note = {R package version 1.40}, url = {https://yihui.org/knitr/}, }

$lme4 @Manual{R-lme4, title = {lme4: Linear Mixed-Effects Models using Eigen and S4}, author = {Douglas Bates and Martin Maechler and Ben Bolker and Steven Walker}, year = {2022}, note = {R package version 1.1-31}, url = {https://github.com/lme4/lme4/}, }

$lmtest @Manual{R-lmtest, title = {lmtest: Testing Linear Regression Models}, author = {Torsten Hothorn and Achim Zeileis and Richard W. Farebrother and Clint Cummins}, year = {2022}, note = {R package version 0.9-40}, url = {https://CRAN.R-project.org/package=lmtest}, }

$lubridate @Manual{R-lubridate, title = {lubridate: Make Dealing with Dates a Little Easier}, author = {Vitalie Spinu and Garrett Grolemund and Hadley Wickham}, year = {2021}, note = {R package version 1.8.0}, url = {https://CRAN.R-project.org/package=lubridate}, }

$maditr @Manual{R-maditr, title = {maditr: Fast Data Aggregation, Modification, and Filtering with Pipes and data.table}, author = {Gregory Demin}, year = {2022}, note = {R package version 0.8.3}, url = {https://github.com/gdemin/maditr}, }

$marginaleffects @Manual{R-marginaleffects, title = {marginaleffects: Marginal Effects, Marginal Means, Predictions, and Contrasts}, author = {Vincent Arel-Bundock}, year = {2022}, note = {R package version 0.8.0}, url = {https://vincentarelbundock.github.io/marginaleffects/}, }

$Matrix @Manual{R-Matrix, title = {Matrix: Sparse and Dense Matrix Classes and Methods}, author = {Douglas Bates and Martin Maechler and Mikael Jagan}, year = {2022}, note = {R package version 1.5-1}, url = {https://CRAN.R-project.org/package=Matrix}, }

$mgcv @Manual{R-mgcv, title = {mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation}, author = {Simon Wood}, year = {2022}, note = {R package version 1.8-41}, url = {https://CRAN.R-project.org/package=mgcv}, }

$modelsummary @Manual{R-modelsummary, title = {modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready}, author = {Vincent Arel-Bundock}, year = {2022}, note = {R package version 1.1.0}, url = {https://vincentarelbundock.github.io/modelsummary/}, }

$nlme @Manual{R-nlme, title = {nlme: Linear and Nonlinear Mixed Effects Models}, author = {José Pinheiro and Douglas Bates and {R Core Team}}, year = {2022}, note = {R package version 3.1-160}, url = {https://svn.r-project.org/R-packages/trunk/nlme/}, }

$psych @Manual{R-psych, title = {psych: Procedures for Psychological, Psychometric, and Personality Research}, author = {William Revelle}, year = {2022}, note = {R package version 2.2.9}, url = {https://personality-project.org/r/psych/ https://personality-project.org/r/psych-manual.pdf}, }

$R.methodsS3 @Manual{R-R.methodsS3, title = {R.methodsS3: S3 Methods Simplified},

author = {Henrik Bengtsson}, year = {2022}, note = {R package version 1.8.2}, url = {https://github.com/HenrikBengtsson/R.methodsS3}, }

$R.oo @Manual{R-R.oo, title = {R.oo: R Object-Oriented Programming with or without References}, author = {Henrik Bengtsson}, year = {2022}, note = {R package version 1.25.0}, url = {https://github.com/HenrikBengtsson/R.oo}, }

$R.utils @Manual{R-R.utils, title = {R.utils: Various Programming Utilities}, author = {Henrik Bengtsson}, year = {2022}, note = {R package version 2.12.1}, url = {https://CRAN.R-project.org/package=R.utils}, }

$readstata13 @Manual{R-readstata13, title = {readstata13: Import Stata Data Files}, author = {Jan Marvin Garbuszus and Sebastian Jeworutzki}, year = {2021}, note = {R package version 0.10.0}, url = {https://github.com/sjewo/readstata13}, }

$readxl @Manual{R-readxl, title = {readxl: Read Excel Files}, author = {Hadley Wickham and Jennifer Bryan}, year = {2022}, note = {R package version 1.4.1}, url = {https://CRAN.R-project.org/package=readxl}, }

$rlang @Manual{R-rlang, title = {rlang: Functions for Base Types and Core R and Tidyverse Features}, author = {Lionel Henry and Hadley Wickham}, year = {2022}, note = {R package version 1.0.6}, url = {https://CRAN.R-project.org/package=rlang}, }

$sandwich @Manual{R-sandwich, title = {sandwich: Robust Covariance Matrix Estimators}, author = {Achim Zeileis and Thomas Lumley}, year = {2022}, note = {R package version 3.0-2}, url = {https://sandwich.R-Forge.R-project.org/}, }

$stargazer @Manual{R-stargazer, title = {stargazer: Well-Formatted Regression and Summary Statistics Tables}, author = {Marek Hlavac}, year = {2022}, note = {R package version 5.2.3}, url = {https://CRAN.R-project.org/package=stargazer}, }

$survival @Manual{R-survival, title = {survival: Survival Analysis}, author = {Terry M Therneau}, year = {2022}, note = {R package version 3.4-0}, url = {https://github.com/therneau/survival}, }

$zoo @Manual{R-zoo, title = {zoo: S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations)}, author = {Achim Zeileis and Gabor Grothendieck and Jeffrey A. Ryan}, year = {2022}, note = {R package version 1.8-11}, url = {https://zoo.R-Forge.R-project.org/}, }

[[34]] @Book{AER2008, title = {Applied Econometrics with {R}}, author = {Christian Kleiber and Achim Zeileis}, year = {2008}, publisher = {Springer-Verlag}, address = {New York}, note = {{ISBN} 978-0-387-77316-2}, url = {https://CRAN.R-project.org/package=AER}, }

[[35]] @Book{car2019, title = {An {R} Companion to Applied Regression}, edition = {Third}, author = {John Fox and Sanford Weisberg}, year = {2019}, publisher = {Sage}, address = {Thousand Oaks {CA}}, url = {https://socialsciences.mcmaster.ca/jfox/Books/Companion/}, }

[[36]] @Article{dagitty2016, title = {Robust causal inference using directed acyclic graphs: the R package 'dagitty'}, journal = {International Journal of Epidemiology}, author =

{Johannes Textor and Benito {van der Zander} and Mark S Gilthorpe and Maciej Liśkiewicz and George TH Ellison}, volume = {45}, number = {6}, pages = {1887–1894}, year = {2016}, doi = {10.1093/ije/dyw341}, }

[[37]] @Article{GJRM2017a, title = {Penalized Likelihood Estimation of a Trivariate Additive Probit Model}, journal = {Biostatistics}, year = {2017}, author = {Filippou P and Marra G and Radice R}, }

[[38]] @Article{GJRM2017b, title = {Bivariate Copula Additive Models for Location, Scale and Shape}, journal = {Computational Statistics and Data Analysis}, year = {2017}, author = {Marra G and Radice R}, }

[[39]] @Article{GJRM2017c, title = {A Simultaneous Equation Approach to Estimating HIV Prevalence with Non-Ignorable Missing Responses}, journal = {Journal of the American Statistical Association}, year = {2017}, author = {Marra G and Radice R and Barnighausen T and Wood SN and McGovern ME}, }

[[40]] @Article{GJRM2016a, title = {Semi-Parametric Copula Sample Selection Models for Count Responses}, journal = {Computational Statistics and Data Analysis}, year = {2016}, author = {Marra G and Wyszynski K}, }

[[41]] @Article{GJRM2016b, title = {Copula Regression Spline Models for Binary Outcomes}, journal = {Statistics and Computing}, year = {2016}, author = {Radice R and Marra G and Wojtys M}, }

[[42]] @Article{GJRM2017d, title = {Testing the Hypothesis of Exogeneity in Regression Spline Bivariate Probit Models}, journal = {Communications in Statistics - Simulation and Computation}, year = {2017}, author = {Marra G and Radice R and Filippou P}, }

[[43]] @Article{GJRM2016c, title = {Bankruptcy Prediction of Small and Medium Enterprises Using a Flexible Binary Generalized Extreme Value Model}, journal = {Journal of the Operational Research Society}, year = {2016}, author = {Calabrese R and Marra G and Osmetti SA}, }

[[44]] @Article{GJRM2013a, title = {A Penalized Likelihood Estimation Approach to Semiparametric Sample Selection Binary Response Modeling}, journal = {Electronic Journal of Statistics}, year = {2013}, author = {Marra G and Radice R}, }

[[45]] @Article{GJRM2013b, title = {Estimation of a Regression Spline Sample Selection Model }, journal = {Computational Statistics and Data Analysis}, year = {2013}, author = {Marra G and Radice R}, }

[[46]] @Article{GJRM2013c, title = {Estimation of a Regression Spline Sample Selection Model }, journal = {Computational Statistics and Data Analysis}, year = {2013}, author = {Marra G and Radice R}, }

[[47]] @Article{GJRM2012a, title = {Coverage Properties of Confidence Intervals for Generalized Additive Model Components}, journal = {Scandinavian Journal of Statistics}, year = {2012}, author = {Marra G and Wood SN}, }

[[48]] @Article{GJRM2011a, title = {Estimation of a Semiparametric Recursive Bivariate Probit Model in the Presence of Endogeneity}, journal = {Canadian Journal of Statistics},

year = {2011}, author = {Marra G and Radice R}, }

[[49]] @Book{knitr2015, title = {Dynamic Documents with {R} and knitr}, author = {Yihui Xie}, publisher = {Chapman and Hall/CRC}, address = {Boca Raton, Florida}, year = {2015}, edition = {2nd}, note = {ISBN 978-1498716963}, url = {https://yihui.org/knitr/}, }

[[50]] @InCollection{knitr2014, booktitle = {Implementing Reproducible Computational Research}, editor = {Victoria Stodden and Friedrich Leisch and Roger D. Peng}, title = {knitr: A Comprehensive Tool for Reproducible Research in {R}}, author = {Yihui Xie}, publisher = {Chapman and Hall/CRC}, year = {2014}, note = {ISBN 978-1466561595}, url = {http://www.crcpress.com/product/isbn/9781466561595}, }

[[51]] @Article{lme42015, title = {Fitting Linear Mixed-Effects Models Using {lme4}}, author = {Douglas Bates and Martin M{"a}chler and Ben Bolker and Steve Walker}, journal = {Journal of Statistical Software}, year = {2015}, volume = {67}, number = {1}, pages = {1–48}, doi = {10.18637/jss.v067.i01}, }

[[52]] @Article{lmtest2002, title = {Diagnostic Checking in Regression Relationships}, author = {Achim Zeileis and Torsten Hothorn}, journal = {R News}, year = {2002}, volume = {2}, number = {3}, pages = {7–10}, url = {https://CRAN.R-project.org/doc/Rnews/}, }

[[53]] @Article{lubridate2011, title = {Dates and Times Made Easy with {lubridate}}, author = {Garrett Grolemund and Hadley Wickham}, journal = {Journal of Statistical Software}, year = {2011}, volume = {40}, number = {3}, pages = {1–25}, url = {https://www.jstatsoft.org/v40/i03/}, }

[[54]] @Article{mgcv2011, title = {Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models}, journal = {Journal of the Royal Statistical Society (B)}, volume = {73}, number = {1}, pages = {3-36}, year = {2011}, author = {S. N. Wood}, }

[[55]] @Article{mgcv2016, title = {Smoothing parameter and model selection for general smooth models (with discussion)}, author = {S.N. Wood and {N.} and {Pya} and B. S{“a}fken}, journal = {Journal of the American Statistical Association}, year = {2016}, pages = {1548-1575}, volume = {111}, }

[[56]] @Article{mgcv2004, title = {Stable and efficient multiple smoothing parameter estimation for generalized additive models}, journal = {Journal of the American Statistical Association}, volume = {99}, number = {467}, pages = {673-686}, year = {2004}, author = {S. N. Wood}, }

[[57]] @Book{mgcv2017, title = {Generalized Additive Models: An Introduction with R}, year = {2017}, author = {S.N Wood}, edition = {2}, publisher = {Chapman and Hall/CRC}, }

[[58]] @Article{mgcv2003, title = {Thin-plate regression splines}, journal = {Journal of the Royal Statistical Society (B)}, volume = {65}, number = {1}, pages = {95-114}, year = {2003}, author = {S. N. Wood}, }

[[59]] @Article{modelsummary2022, title = {{modelsummary}: Data and Model Summaries in {R}}, author = {Vincent Arel-Bundock}, journal = {Journal of Statistical Software}, year = {2022}, volume = {103}, number = {1}, pages = {1–23}, doi = {10.18637/jss.v103.i01}, }

[[60]] @Book{nlme2000, title = {Mixed-Effects Models in S and S-PLUS}, author = {José C. Pinheiro and Douglas M. Bates}, year = {2000}, publisher = {Springer}, address = {New York}, doi = {10.1007/b98882}, }

[[61]] @InProceedings{R.methodsS32003, author = {Henrik Bengtsson}, title = {The {R.oo} package - Object-Oriented Programming with References Using Standard {R} Code}, booktitle = {Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)}, year = {2003}, editor = {Kurt Hornik and Friedrich Leisch and Achim Zeileis}, address = {Vienna, Austria}, month = {March}, issn = {1609-395X}, url = {https://www.r-project.org/conferences/DSC-2003/Proceedings/Bengtsson.pdf}, howpublished = {https://www.r-project.org/conferences/DSC-2003/Proceedings/}, }

[[62]] @InProceedings{R.oo2003, author = {Henrik Bengtsson}, title = {The {R.oo} package - Object-Oriented Programming with References Using Standard {R} Code}, booktitle = {Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)}, year = {2003}, editor = {Kurt Hornik and Friedrich Leisch and Achim Zeileis}, address = {Vienna, Austria}, month = {March}, issn = {1609-395X}, url = {https://www.r-project.org/conferences/DSC-2003/Proceedings/Bengtsson.pdf}, howpublished = {https://www.r-project.org/conferences/DSC-2003/Proceedings/}, }

[[63]] @Article{sandwich2020, title = {Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in {R}}, author = {Achim Zeileis and Susanne K"oll and Nathaniel Graham}, journal = {Journal of Statistical Software}, year = {2020}, volume = {95}, number = {1}, pages = {1–36}, doi = {10.18637/jss.v095.i01}, }

[[64]] @Article{sandwich2004, title = {Econometric Computing with {HC} and {HAC} Covariance Matrix Estimators}, author = {Achim Zeileis}, journal = {Journal of Statistical Software}, year = {2004}, volume = {11}, number = {10}, pages = {1–17}, doi = {10.18637/jss.v011.i10}, }

[[65]] @Article{sandwich2006, title = {Object-Oriented Computation of Sandwich Estimators}, author = {Achim Zeileis}, journal = {Journal of Statistical Software}, year = {2006}, volume = {16}, number = {9}, pages = {1–16}, doi = {10.18637/jss.v016.i09}, }

[[66]] @Book{survival-book, title = {Modeling Survival Data: Extending the {C}ox Model}, author = {{Terry M. Therneau} and {Patricia M. Grambsch}}, year = {2000}, publisher = {Springer}, address = {New York}, isbn = {0-387-98784-3}, }

[[67]] @Article{zoo2005, title = {zoo: S3 Infrastructure for Regular and Irregular Time Series}, author = {Achim Zeileis and Gabor Grothendieck}, journal = {Journal of Statistical Software}, year = {2005}, volume = {14}, number = {6}, pages = {1–27}, doi = {10.18637/jss.v014.i06}, }