ARE 213 PS 1b

S. Sung, H. Husain, T. Woolley, A. Watt

2021-10-18

Contents

Packages

```
# install.packages("pacman")
# install.packages("tidyverse")
# install.packages("plm")
# install.packages('foreign')
# install.packages('stargazer')
# install.packages("finalfit")
```

```r
# install.packages("glmnet")
# install.packages("jtools")
# install.packages("Hmisc")
library(tidyverse)
library(foreign)
# library(xtable)
library(stargazer)
library(finalfit)
library(glmnet)
library(jtools)  # summ() for regression summaries
```

Data cleaning from PS 1a

```r
data = read.dta('ps1.dta')
missing_codes = read.csv('missing_codes.csv')
mvars = as.character(missing_codes$varname)
missing_codes$num_missing = as.integer(0)
for (row in 1:nrow(missing_codes)) {
  var = as.character(missing_codes[row, "varname"])
  code  = as.numeric(missing_codes[row, "missing_code"])
  nmissing = as.integer(sum(data[, var] == code))
  missing_codes$num_missing[missing_codes$varname==var] = nmissing
  data[, var] = na_if(data[, var], code)
}
# Convert all variables with <7 unique values to factor (and 3 additional variables)
factor_vars = c("isllb10", "birmon", "weekday")
for (var in colnames(data)) {
  if (length(unique(data[!is.na(data[, var]), var])) < 7 || var %in% factor_vars) {
    data[, var] = factor(data[, var])
  }
}
# label data
variable_labels_df = read.csv('variable_labels.csv')
variable_labels <- setNames(as.character(variable_labels_df$label), variable_labels_df$varname)
data <- Hmisc::upData(data, labels = variable_labels)
```

```
## Input object size:    30868856 bytes;     48 variables    120461 observations
## New object size: 30886992 bytes; 48 variables    120461 observations
```

```r
# Dataframe with missing dropped
df = data[complete.cases(data), ]
```

Problem 1

In Problem Set 1a, you used linear regression to relate infant health outcomes and maternal smoking during pregnancy.

Part (a)

Under the assumption of random assignment conditional on the observables, what are the sources of misspecification bias in the estimates generated by the linear model estimated in Problem Set 1a?

- functional form: what if the CEF isn't linear in parameters. Even if we have a linear CEF, what if we didn't include important interaction terms?

Part (b)

Now, consider a series estimator. Estimate the smoking effects using a flexible functional form for the control variables (e.g., higher order terms and interactions). What are the benefits and drawbacks to this approach?

```
df1b = df %>% select(dbrwt, tobacco, csex, mrace3, preterm,
                     dmage, dfage, dmeduc, dfeduc, ormoth, orfath,
                     disllb, dtotord, dmar, adequacy, nprevist)

# indicator vars (no higher order terms)
vars1 = names(Filter(is.factor, select(df1b, -dbrwt)))
# quantitative vars (need to create higher order terms)
vars2 = names(Filter(is.integer, select(df1b, -dbrwt)))

birthweight = df1b$dbrwt
x = df1b %>% select(-dbrwt)
# Create dummies from factor variables, all interactions, and squared continuous vars
formula1 = as.formula(paste("~ .^2 +", paste0("I(",vars2,"^2)", collapse=' + ') ))
xx <- model.matrix(formula1, x)[, -1]

# Series Regression
reg_1b = lm(birthweight ~ xx)
# rename the coefficients
names(reg_1b$coefficients) <- gsub("xx","",names(reg_1b$coefficients))
coefficients(reg_1b)
```

```
##        (Intercept)          tobacco2             csex2           mrace32
##       9.888205e+02      2.144189e+02     -2.460893e+01     -2.149490e+02
##            mrace33          preterm2             dmage             dfage
##      -1.219773e+02      8.118905e+02      4.227047e+01     -5.473113e+00
##             dmeduc            dfeduc           ormoth1           ormoth2
##       1.791153e+01     -1.928443e+01      3.887857e+02      6.658417e+01
##            ormoth3           ormoth4           ormoth5           orfath1
##      -9.894276e+02      3.895786e+02     -1.888306e+02     -6.865286e+02
##            orfath2           orfath3           orfath4           orfath5
##      -1.378863e+02      3.982333e+02      1.606181e+02      4.247760e+02
##             disllb           dtotord             dmar2         adequacy2
##      -7.281393e-02     -4.211551e+01     -2.200343e+02      5.923781e+02
##          adequacy3          nprevist        I(dmage^2)        I(dfage^2)
##       1.047928e+03      1.661254e+02     -9.170697e-01     -1.479438e-02
##        I(dmeduc^2)       I(dfeduc^2)       I(disllb^2)      I(dtotord^2)
##      -3.130084e-02      4.486755e-01     -9.403191e-04     -1.545961e+00
##      I(nprevist^2)     tobacco2:csex2  tobacco2:mrace32  tobacco2:mrace33
##      -3.866891e+00      1.851477e+00     -4.178789e+01     -6.325021e+01
##   tobacco2:preterm2    tobacco2:dmage    tobacco2:dfage   tobacco2:dmeduc
##      -1.402385e+00      4.805547e+00      1.393528e+00     -1.100485e+01
##     tobacco2:dfeduc  tobacco2:ormoth1  tobacco2:ormoth2  tobacco2:ormoth3
##      -2.502236e+00     -6.168336e+01      2.487877e+00     -3.678457e+02
##    tobacco2:ormoth4  tobacco2:ormoth5  tobacco2:orfath1  tobacco2:orfath2
##      -1.700206e+02     -4.560742e+01      2.174027e+02     -8.900712e+01
##    tobacco2:orfath3  tobacco2:orfath4  tobacco2:orfath5   tobacco2:disllb
```

```
##       -2.287477e+02       -2.542848e+02       -6.681166e+01       -1.951305e-02
##    tobacco2:dtotord      tobacco2:dmar2 tobacco2:adequacy2 tobacco2:adequacy3
##        8.154257e+00       -3.697862e+01        3.134864e+01        2.348893e+01
##   tobacco2:nprevist        csex2:mrace32        csex2:mrace33      csex2:preterm2
##        4.642094e-02        2.656400e+01        1.084581e+01       -1.366212e+01
##          csex2:dmage         csex2:dfage        csex2:dmeduc        csex2:dfeduc
##        1.317171e-01       -1.104886e+00        2.759663e+00       -3.474496e+00
##       csex2:ormoth1       csex2:ormoth2       csex2:ormoth3       csex2:ormoth4
##        1.181547e+02        1.787680e+01       -2.727092e+01       -3.465208e+00
##       csex2:ormoth5       csex2:orfath1       csex2:orfath2       csex2:orfath3
##        1.112386e+02       -1.454543e+02       -2.383637e+01       -8.227054e+00
##       csex2:orfath4       csex2:orfath5        csex2:disllb       csex2:dtotord
##        7.675028e+01       -3.865138e+01        1.865434e-02       -3.701548e+00
##         csex2:dmar2    csex2:adequacy2    csex2:adequacy3      csex2:nprevist
##       -4.395974e+00       -1.151253e+01       -3.651453e+01       -3.984337e+00
##    mrace32:preterm2    mrace33:preterm2        mrace32:dmage        mrace33:dmage
##        6.161528e+01        2.848803e+01        7.125760e+00       -1.048850e+00
##       mrace32:dfage       mrace33:dfage      mrace32:dmeduc      mrace33:dmeduc
##       -3.940058e+00       -8.698864e-01       -3.095915e+00       -6.617296e-01
##      mrace32:dfeduc      mrace33:dfeduc     mrace32:ormoth1     mrace33:ormoth1
##        1.284459e+00       -8.763940e+00        1.216725e+02        2.744671e+02
##     mrace32:ormoth2     mrace33:ormoth2     mrace32:ormoth3     mrace33:ormoth3
##        1.049448e+02        1.920420e+02       -4.502423e+00       -2.517420e+01
##     mrace32:ormoth4     mrace33:ormoth4     mrace32:ormoth5     mrace33:ormoth5
##        2.748790e+02        2.577802e+01        3.673992e+02       -3.778896e+00
##     mrace32:orfath1     mrace33:orfath1     mrace32:orfath2     mrace33:orfath2
##       -7.637054e+01       -1.939242e+01        1.637044e+02        2.002011e+01
##     mrace32:orfath3     mrace33:orfath3     mrace32:orfath4     mrace33:orfath4
##                  NA        5.410640e+02        2.297409e+02        7.550759e+01
##     mrace32:orfath5     mrace33:orfath5      mrace32:disllb      mrace33:disllb
##       -3.312549e+02        1.081440e+02        5.199514e-02       -4.686480e-03
##     mrace32:dtotord     mrace33:dtotord       mrace32:dmar2       mrace33:dmar2
##        2.256630e-02       -8.223725e+00        8.361100e+01       -2.498313e+01
##   mrace32:adequacy2  mrace33:adequacy2  mrace32:adequacy3  mrace33:adequacy3
##       -6.314167e+01        2.847886e+01       -1.900216e+01        1.052331e+02
##    mrace32:nprevist    mrace33:nprevist      preterm2:dmage      preterm2:dfage
##       -8.383459e+00        1.324633e+01       -3.405302e+00       -2.180616e+00
##     preterm2:dmeduc     preterm2:dfeduc    preterm2:ormoth1    preterm2:ormoth2
##        1.467995e+00        6.611429e+00       -3.996192e+02       -1.929205e+02
##    preterm2:ormoth3    preterm2:ormoth4    preterm2:ormoth5    preterm2:orfath1
##                  NA        1.051697e+02        1.082922e+02        8.902877e+02
##    preterm2:orfath2    preterm2:orfath3    preterm2:orfath4    preterm2:orfath5
##        2.023559e+02                  NA       -3.941754e+02       -6.667618e+01
##     preterm2:disllb    preterm2:dtotord      preterm2:dmar2 preterm2:adequacy2
##        5.292998e-01       -6.075104e+00       -1.006683e+01       -1.141571e+02
## preterm2:adequacy3   preterm2:nprevist         dmage:dfage        dmage:dmeduc
##       -2.953670e+02       -2.527863e+01        1.313347e-01        3.417724e-01
##        dmage:dfeduc       dmage:ormoth1       dmage:ormoth2       dmage:ormoth3
##        1.902418e-02        5.871341e+00        2.501944e-01        2.285636e+01
##       dmage:ormoth4       dmage:ormoth5       dmage:orfath1       dmage:orfath2
##       -4.765782e+00        1.137417e+00       -8.048578e+00        3.588383e+00
##       dmage:orfath3       dmage:orfath4       dmage:orfath5        dmage:disllb
##       -9.368101e+00       -6.200945e+00        4.517460e+00       -6.642264e-03
##       dmage:dtotord         dmage:dmar2     dmage:adequacy2     dmage:adequacy3
```

```
##      1.784992e+00      -5.039936e+00      -1.040993e+00      -3.533706e+00
##     dmage:nprevist       dfage:dmeduc       dfage:dfeduc      dfage:ormoth1
##     -2.812148e-01      -3.166786e-01      -9.461215e-03      -1.943479e-01
##     dfage:ormoth2      dfage:ormoth3      dfage:ormoth4      dfage:ormoth5
##      1.623410e-01      -1.482097e+01       4.803686e+00      -1.390343e+00
##     dfage:orfath1      dfage:orfath2      dfage:orfath3      dfage:orfath4
##     -8.123761e+00      -1.395197e+00       1.949553e+01       1.141395e+01
##     dfage:orfath5       dfage:disllb      dfage:dtotord        dfage:dmar2
##     -8.361689e-01       1.261312e-03       3.723985e-01       7.323969e-01
##   dfage:adequacy2    dfage:adequacy3      dfage:nprevist       dmeduc:dfeduc
##      8.059195e-01       3.166813e+00       5.957992e-01       6.692504e-01
##    dmeduc:ormoth1     dmeduc:ormoth2     dmeduc:ormoth3     dmeduc:ormoth4
##     -1.426528e+01      -2.496920e+01      -4.716511e+01      -1.599738e+01
##    dmeduc:ormoth5     dmeduc:orfath1     dmeduc:orfath2     dmeduc:orfath3
##      2.533409e+01       1.837127e+01       1.645496e+00      -1.517077e+01
##    dmeduc:orfath4     dmeduc:orfath5      dmeduc:disllb     dmeduc:dtotord
##      1.244237e+01      -4.575611e+00       4.616856e-03      -5.572140e-01
##      dmeduc:dmar2   dmeduc:adequacy2   dmeduc:adequacy3    dmeduc:nprevist
##      2.678595e+00      -6.460783e+00      -1.086803e+01      -1.150319e+00
##    dfeduc:ormoth1     dfeduc:ormoth2     dfeduc:ormoth3     dfeduc:ormoth4
##      3.283190e+00       2.397848e+01       5.879949e+01       3.921535e+00
##    dfeduc:ormoth5     dfeduc:orfath1     dfeduc:orfath2     dfeduc:orfath3
##     -5.534116e+00      -2.120640e+01      -8.056381e+00       2.675163e+00
##    dfeduc:orfath4     dfeduc:orfath5      dfeduc:disllb     dfeduc:dtotord
##     -3.093351e-01      -2.033237e+00      -2.503926e-03      -6.272049e-01
##      dfeduc:dmar2   dfeduc:adequacy2   dfeduc:adequacy3    dfeduc:nprevist
##      1.253461e+01       2.335086e+00       7.385599e+00      -1.971177e-01
##   ormoth1:orfath1    ormoth2:orfath1    ormoth3:orfath1    ormoth4:orfath1
##     -1.776174e+02      -3.307756e+01       5.804182e+02      -7.116497e+01
##   ormoth5:orfath1    ormoth1:orfath2    ormoth2:orfath2    ormoth3:orfath2
##     -4.135189e+02       1.687862e+02      -2.319910e+01       4.438676e+02
##   ormoth4:orfath2    ormoth5:orfath2    ormoth1:orfath3    ormoth2:orfath3
##      9.760434e+01      -1.592669e+01                 NA      -2.064433e+02
##   ormoth3:orfath3    ormoth4:orfath3    ormoth5:orfath3    ormoth1:orfath4
##     -1.142997e+02      -6.104628e+02                 NA      -1.053158e+02
##   ormoth2:orfath4    ormoth3:orfath4    ormoth4:orfath4    ormoth5:orfath4
##      3.191826e+01       2.946576e+02       6.953357e+01       4.612715e+01
##   ormoth1:orfath5    ormoth2:orfath5    ormoth3:orfath5    ormoth4:orfath5
##     -3.547698e+02      -5.914413e+01      -4.682794e+02      -2.481097e+02
##   ormoth5:orfath5     ormoth1:disllb     ormoth2:disllb     ormoth3:disllb
##     -4.735882e+01      -1.286068e-01       2.383762e-02       2.653143e-01
##    ormoth4:disllb     ormoth5:disllb    ormoth1:dtotord    ormoth2:dtotord
##     -7.031207e-02       1.232049e-01      -1.350182e+01       2.198941e+01
##   ormoth3:dtotord    ormoth4:dtotord    ormoth5:dtotord      ormoth1:dmar2
##      1.376697e+02      -4.039166e+01      -8.720333e+00       8.158766e+01
##     ormoth2:dmar2      ormoth3:dmar2      ormoth4:dmar2      ormoth5:dmar2
##     -5.239427e+01       2.071374e+02       1.640844e+01       2.835995e+01
## ormoth1:adequacy2  ormoth2:adequacy2  ormoth3:adequacy2  ormoth4:adequacy2
##     -8.548134e+01      -2.450428e+01      -7.218578e+01      -9.999306e+01
## ormoth5:adequacy2  ormoth1:adequacy3  ormoth2:adequacy3  ormoth3:adequacy3
##     -7.364570e+01       1.450182e+02      -4.475147e+01       7.405109e+02
## ormoth4:adequacy3  ormoth5:adequacy3    ormoth1:nprevist   ormoth2:nprevist
##      5.201581e+01      -2.032143e+02       9.073751e+00      -1.146680e+00
##  ormoth3:nprevist   ormoth4:nprevist    ormoth5:nprevist     orfath1:disllb
```

```
##        3.910174e+01       -1.107515e+01       -2.285245e+01        1.174103e-01
##      orfath2:disllb      orfath3:disllb      orfath4:disllb      orfath5:disllb
##       -8.585136e-02        5.397238e-02       -3.397815e-02       -9.326690e-02
##     orfath1:dtotord     orfath2:dtotord     orfath3:dtotord     orfath4:dtotord
##        4.550441e+01       -3.079793e+01       -2.412029e+01       -1.057973e+01
##     orfath5:dtotord       orfath1:dmar2       orfath2:dmar2       orfath3:dmar2
##       -4.379531e+01        8.841007e+01        2.853187e+01        2.855421e+02
##       orfath4:dmar2       orfath5:dmar2  orfath1:adequacy2  orfath2:adequacy2
##        3.667375e+01       -3.291933e+01        6.500342e+01        6.563400e+01
##   orfath3:adequacy2  orfath4:adequacy2  orfath5:adequacy2  orfath1:adequacy3
##        2.997188e+01        1.756439e+01       -2.688476e+01       -2.040077e+01
##   orfath2:adequacy3  orfath3:adequacy3  orfath4:adequacy3  orfath5:adequacy3
##        1.250236e+02       -1.092874e+03        8.306446e+01       -1.448977e+02
##    orfath1:nprevist    orfath2:nprevist    orfath3:nprevist    orfath4:nprevist
##       -7.475106e+00        8.818509e+00       -1.816355e+01        3.252319e+00
##    orfath5:nprevist      disllb:dtotord        disllb:dmar2   disllb:adequacy2
##       -1.258824e+01       -2.604213e-02        2.020447e-02        7.267905e-02
##    disllb:adequacy3     disllb:nprevist       dtotord:dmar2  dtotord:adequacy2
##        1.986484e-01        2.369469e-02        8.046503e+00        2.098157e+01
##   dtotord:adequacy3    dtotord:nprevist     dmar2:adequacy2    dmar2:adequacy3
##        2.735461e+01        5.057545e-01       -8.390896e+00        3.123210e+01
##      dmar2:nprevist adequacy2:nprevist adequacy3:nprevist
##        1.034898e+01       -3.358381e+01       -6.527302e+01
```

Need to change to longtable before final submission

```
A table enviroment cannot be broken across pages. Delete \begin{table}\centering and \end{table}, replac
```

```
stargazer(reg_1b, title="Series Regression", header=FALSE, single.row=TRUE,
          se = NULL, notes = 'SEs omitted for brevity',  type ="html", report = "vc*")  # , align = TRU
```

Series Regression

Dependent variable:

birthweight

tobacco2

214.419***

csex2

-24.609

mrace32

-214.949

mrace33

-121.977

preterm2

811.890***

dmage

42.270***

dfage

-5.473

dmeduc

17.912

dfeduc

-19.284

ormoth1

388.786

ormoth2

66.584

ormoth3

-989.428

ormoth4

389.579

ormoth5

-188.831

orfath1

-686.529

orfath2

-137.886

orfath3

398.233

orfath4

160.618

orfath5

424.776

disllb

-0.073

dtotord

-42.116**

dmar2

-220.034***

adequacy2

592.378***

adequacy3

1,047.928***

nprevist

166.125***

I(dmage2)

-0.917***

I(dfage2)

-0.015

I(dmeduc2)

-0.031

I(dfeduc2)

0.449

I(disllb2)

-0.001***

I(dtotord2)

-1.546***

I(nprevist2)

-3.867***

tobacco2:csex2

1.851

tobacco2:mrace32

-41.788

tobacco2:mrace33

-63.250***

tobacco2:preterm2

-1.402

tobacco2:dmage

4.806***

tobacco2:dfage

1.394

tobacco2:dmeduc

-11.005***

tobacco2:dfeduc

-2.502

tobacco2:ormoth1

-61.683

tobacco2:ormoth2

2.488

tobacco2:ormoth3

-367.846

tobacco2:ormoth4

-170.021

tobacco2:ormoth5

-45.607

tobacco2:orfath1

217.403**

tobacco2:orfath2

-89.007**

tobacco2:orfath3

-228.748

tobacco2:orfath4

-254.285**

tobacco2:orfath5

-66.812

tobacco2:disllb

-0.020

tobacco2:dtotord

8.154**

tobacco2:dmar2

-36.979***

tobacco2:adequacy2

31.349**

tobacco2:adequacy3

23.489

tobacco2:nprevist

0.046

csex2:mrace32

26.564

csex2:mrace33

10.846

csex2:preterm2

-13.662

csex2:dmage

0.132

csex2:dfage

-1.105

csex2:dmeduc

2.760

csex2:dfeduc

-3.474*

csex2:ormoth1

118.155*

csex2:ormoth2

17.877

csex2:ormoth3

-27.271

csex2:ormoth4

-3.465

csex2:ormoth5

111.239**

csex2:orfath1

-145.454**

csex2:orfath2

-23.836

csex2:orfath3

-8.227

csex2:orfath4

76.750

csex2:orfath5

-38.651

csex2:disllb

0.019

csex2:dtotord

-3.702

csex2:dmar2

-4.396

csex2:adequacy2

-11.513

csex2:adequacy3

-36.515*

csex2:nprevist

-3.984***

mrace32:preterm2

61.615

mrace33:preterm2

28.488

mrace32:dmage

7.126**

mrace33:dmage

-1.049

mrace32:dfage

-3.940

mrace33:dfage

-0.870

mrace32:dmeduc

-3.096

mrace33:dmeduc

-0.662

mrace32:dfeduc

1.284

mrace33:dfeduc

-8.764**

mrace32:ormoth1

121.673

mrace33:ormoth1

274.467

mrace32:ormoth2

104.945

mrace33:ormoth2

192.042***

mrace32:ormoth3

-4.502

mrace33:ormoth3

-25.174

mrace32:ormoth4

274.879

mrace33:ormoth4

25.778

mrace32:ormoth5

367.399*

mrace33:ormoth5

-3.779

mrace32:orfath1

-76.371

mrace33:orfath1

-19.392

mrace32:orfath2

163.704

mrace33:orfath2

20.020

mrace32:orfath3

mrace33:orfath3

541.064*

mrace32:orfath4

229.741

mrace33:orfath4

75.508

mrace32:orfath5

-331.255

mrace33:orfath5

108.144

mrace32:disllb

0.052

mrace33:disllb

-0.005

mrace32:dtotord

0.023

mrace33:dtotord

-8.224*

mrace32:dmar2

83.611*

mrace33:dmar2

-24.983*

mrace32:adequacy2

-63.142*

mrace33:adequacy2

28.479*

mrace32:adequacy3

-19.002

mrace33:adequacy3

105.233***

mrace32:nprevist

-8.383*

mrace33:nprevist

13.246***

preterm2:dmage

-3.405

preterm2:dfage

-2.181

preterm2:dmeduc

1.468

preterm2:dfeduc

6.611

preterm2:ormoth1

-399.619

preterm2:ormoth2

-192.920**

preterm2:ormoth3

preterm2:ormoth4

105.170

preterm2:ormoth5

108.292

preterm2:orfath1

890.288***

preterm2:orfath2

202.356**

preterm2:orfath3

preterm2:orfath4

-394.175*

preterm2:orfath5

-66.676

preterm2:disllb

0.529

preterm2:dtotord

-6.075

preterm2:dmar2

-10.067

preterm2:adequacy2

-114.157***

preterm2:adequacy3

-295.367***

preterm2:nprevist

-25.279***

dmage:dfage

0.131

dmage:dmeduc

0.342

dmage:dfeduc

0.019

dmage:ormoth1

5.871

dmage:ormoth2

0.250

dmage:ormoth3

22.856

dmage:ormoth4

-4.766

dmage:ormoth5

1.137

dmage:orfath1

-8.049

dmage:orfath2

3.588

dmage:orfath3

-9.368

dmage:orfath4

-6.201

dmage:orfath5

4.517

dmage:disllb

-0.007***

dmage:dtotord

1.785***

dmage:dmar2

-5.040***

dmage:adequacy2

-1.041

dmage:adequacy3

-3.534

dmage:nprevist

-0.281

dfage:dmeduc

-0.317

dfage:dfeduc

-0.009

dfage:ormoth1

-0.194

dfage:ormoth2

0.162

dfage:ormoth3

-14.821

dfage:ormoth4

4.804

dfage:ormoth5

-1.390

dfage:orfath1

-8.124

dfage:orfath2

-1.395

dfage:orfath3

19.496

dfage:orfath4

11.414*

dfage:orfath5

-0.836

dfage:disllb

0.001

dfage:dtotord

0.372

dfage:dmar2

0.732

dfage:adequacy2

0.806

dfage:adequacy3

3.167

dfage:nprevist

0.596***

dmeduc:dfeduc

0.669

dmeduc:ormoth1

-14.265

dmeduc:ormoth2

-24.969***

dmeduc:ormoth3

-47.165

dmeduc:ormoth4

-15.997

dmeduc:ormoth5

25.334

dmeduc:orfath1

18.371

dmeduc:orfath2

1.645

dmeduc:orfath3

-15.171

dmeduc:orfath4

12.442

dmeduc:orfath5

-4.576

dmeduc:disllb

0.005

dmeduc:dtotord

-0.557

dmeduc:dmar2

2.679

dmeduc:adequacy2

-6.461**

dmeduc:adequacy3

-10.868*

dmeduc:nprevist

-1.150***

dfeduc:ormoth1

3.283

dfeduc:ormoth2

23.978***

dfeduc:ormoth3

58.799

dfeduc:ormoth4

3.922

dfeduc:ormoth5

-5.534

dfeduc:orfath1

-21.206

dfeduc:orfath2

-8.056

dfeduc:orfath3

2.675

dfeduc:orfath4

-0.309

dfeduc:orfath5

-2.033

dfeduc:disllb

-0.003

dfeduc:dtotord

-0.627

dfeduc:dmar2

12.535***

dfeduc:adequacy2

2.335

dfeduc:adequacy3

7.386

dfeduc:nprevist

-0.197

ormoth1:orfath1

-177.617*

ormoth2:orfath1

-33.078

ormoth3:orfath1

580.418

ormoth4:orfath1

-71.165

ormoth5:orfath1

-413.519

ormoth1:orfath2

168.786

ormoth2:orfath2

-23.199

ormoth3:orfath2

443.868

ormoth4:orfath2

97.604

ormoth5:orfath2

-15.927

ormoth1:orfath3

ormoth2:orfath3

-206.443

ormoth3:orfath3

-114.300

ormoth4:orfath3

-610.463*

ormoth5:orfath3

ormoth1:orfath4

-105.316

ormoth2:orfath4

31.918

ormoth3:orfath4

294.658

ormoth4:orfath4

69.534

ormoth5:orfath4

46.127

ormoth1:orfath5

-354.770

ormoth2:orfath5

-59.144

ormoth3:orfath5

-468.279

ormoth4:orfath5

-248.110

ormoth5:orfath5

-47.359

ormoth1:disllb

-0.129

ormoth2:disllb

0.024

ormoth3:disllb

0.265

ormoth4:disllb

-0.070

ormoth5:disllb

0.123

ormoth1:dtotord

-13.502

ormoth2:dtotord

21.989

ormoth3:dtotord

137.670**

ormoth4:dtotord

-40.392

ormoth5:dtotord

-8.720

ormoth1:dmar2

81.588

ormoth2:dmar2

-52.394

ormoth3:dmar2

207.137

ormoth4:dmar2

16.408

ormoth5:dmar2

28.360

ormoth1:adequacy2

-85.481

ormoth2:adequacy2

-24.504

ormoth3:adequacy2

-72.186

ormoth4:adequacy2

-99.993

ormoth5:adequacy2

-73.646

ormoth1:adequacy3

145.018

ormoth2:adequacy3

-44.751

ormoth3:adequacy3

740.511

ormoth4:adequacy3

52.016

ormoth5:adequacy3

-203.214

ormoth1:nprevist

9.074

ormoth2:nprevist

-1.147

ormoth3:nprevist

39.102

ormoth4:nprevist

-11.075

ormoth5:nprevist

-22.852***

orfath1:disllb

0.117

orfath2:disllb

-0.086

orfath3:disllb

0.054

orfath4:disllb

-0.034

orfath5:disllb

-0.093

orfath1:dtotord

45.504*

orfath2:dtotord

-30.798**

orfath3:dtotord

-24.120

orfath4:dtotord

-10.580

orfath5:dtotord

-43.795

orfath1:dmar2

88.410

orfath2:dmar2

28.532

orfath3:dmar2

285.542

orfath4:dmar2

36.674

orfath5:dmar2

-32.919

orfath1:adequacy2

65.003

orfath2:adequacy2

65.634*

orfath3:adequacy2

29.972

orfath4:adequacy2

17.564

orfath5:adequacy2

-26.885

orfath1:adequacy3

-20.401

orfath2:adequacy3

125.024*

orfath3:adequacy3

-1,092.874***

orfath4:adequacy3

83.064

orfath5:adequacy3

-144.898

orfath1:nprevist

-7.475

orfath2:nprevist

8.819*

orfath3:nprevist

-18.164

orfath4:nprevist

3.252

orfath5:nprevist

-12.588

disllb:dtotord

-0.026***

disllb:dmar2

0.020

disllb:adequacy2

0.073***

disllb:adequacy3

0.199***

disllb:nprevist

0.024***

dtotord:dmar2

8.047*

dtotord:adequacy2

20.982***

dtotord:adequacy3

27.355***

dtotord:nprevist

0.506

dmar2:adequacy2

-8.391

dmar2:adequacy3

31.232

dmar2:nprevist

10.349***

adequacy2:nprevist

-33.584***

adequacy3:nprevist

-65.273***

Constant

988.820***

Observations

114,610

R2

0.132

Adjusted R2

0.130

Residual Std. Error

545.824 (df = 114304)

F Statistic

57.135*** (df = 305; 114304)

Note:

p<0.1; ***p<0.05;*** p<0.01

SEs omitted for brevity

Part (c)

Use the LASSO to determine which covariates (and higher order terms) to include in your regression from part (b). Do you end up dropping some covariates that you had thought might be necessary to include?

```
# use glmnet with alpha=1 for lasso
reg_1c = glmnet(xx, birthweight, family="gaussian", alpha=1)
# print results (Df = # of variables, %Dev = R^2)
print(reg_1c)
```

```
##
## Call:  glmnet(x = xx, y = birthweight, family = "gaussian", alpha = 1)
##
##      Df  %Dev  Lambda
## 1     0  0.00 113.900
## 2     1  0.64 103.800
## 3     1  1.18  94.590
## 4     1  1.62  86.180
## 5     3  2.11  78.530
## 6     4  2.78  71.550
## 7     6  3.37  65.200
## 8     6  4.12  59.400
## 9     7  4.77  54.130
## 10   10  5.38  49.320
## 11   10  6.01  44.940
## 12   10  6.53  40.940
## 13   10  6.96  37.310
## 14   11  7.33  33.990
## 15   12  7.63  30.970
## 16   12  7.90  28.220
## 17   13  8.13  25.710
## 18   14  8.32  23.430
## 19   14  8.51  21.350
## 20   15  8.68  19.450
## 21   16  8.84  17.720
## 22   15  9.00  16.150
## 23   16  9.13  14.710
## 24   17  9.27  13.410
## 25   20  9.38  12.220
## 26   23  9.48  11.130
## 27   29  9.57  10.140
## 28   37  9.77   9.241
## 29   46 10.00   8.420
## 30   47 10.34   7.672
## 31   51 10.64   6.991
## 32   53 10.90   6.370
## 33   50 11.09   5.804
```

```
## 34   53 11.25   5.288
## 35   56 11.38   4.818
## 36   57 11.50   4.390
## 37   60 11.59   4.000
## 38   65 11.68   3.645
## 39   68 11.77   3.321
## 40   70 11.86   3.026
## 41   72 11.93   2.757
## 42   74 12.01   2.512
## 43   77 12.08   2.289
## 44   82 12.14   2.086
## 45   85 12.22   1.900
## 46   92 12.31   1.732
## 47   96 12.39   1.578
## 48  101 12.47   1.438
## 49  105 12.51   1.310
## 50  115 12.56   1.194
## 51  119 12.61   1.088
## 52  122 12.65   0.991
## 53  126 12.70   0.903
## 54  132 12.72   0.823
## 55  140 12.75   0.750
## 56  149 12.79   0.683
## 57  155 12.83   0.622
## 58  159 12.85   0.567
## 59  165 12.88   0.517
## 60  170 12.91   0.471
## 61  183 12.93   0.429
## 62  190 12.96   0.391
## 63  197 12.98   0.356
## 64  200 13.00   0.324
## 65  203 13.02   0.296
## 66  204 13.03   0.269
## 67  211 13.05   0.246
## 68  214 13.06   0.224
## 69  222 13.07   0.204
## 70  222 13.08   0.186
## 71  230 13.09   0.169
## 72  234 13.10   0.154
## 73  240 13.11   0.140
## 74  244 13.12   0.128
## 75  246 13.13   0.117
## 76  251 13.14   0.106
## 77  255 13.15   0.097
## 78  258 13.15   0.088
## 79  261 13.16   0.080
## 80  266 13.16   0.073
## 81  272 13.17   0.067
## 82  272 13.17   0.061
## 83  277 13.18   0.055
## 84  277 13.18   0.050
## 85  277 13.18   0.046
## 86  280 13.19   0.042
## 87  280 13.19   0.038
```

```
## 88  281 13.19   0.035
## 89  282 13.19   0.032
## 90  284 13.19   0.029
## 91  286 13.20   0.026
## 92  290 13.20   0.024
## 93  291 13.20   0.022
## 94  295 13.20   0.020
## 95  297 13.20   0.018
## 96  299 13.20   0.017
## 97  299 13.20   0.015
## 98  300 13.20   0.014
## 99  301 13.20   0.012
## 100 301 13.20   0.011
```

```r
# if we wanted to limit the model to 20 variables, the 25th iteration
#   where lambda = 12.220 gives 20 variables (and decreases in )
choice = 25
lambda = reg_1c$lambda[[choice]]
print(paste('# of variables in 25th iteration:', sum(reg_1c$beta[, choice] != 0)))
```

```
## [1] "# of variables in 25th iteration: 20"
```

```r
# print the 25th lasso regression coefficients
print(reg_1c$beta[, choice])
```

```
##          tobacco2             csex2            mrace32            mrace33
##      0.000000e+00     -1.004847e+02      -6.707636e+01      -1.645689e+02
##          preterm2             dmage              dfage             dmeduc
##      0.000000e+00      0.000000e+00       0.000000e+00       0.000000e+00
##            dfeduc           ormoth1            ormoth2            ormoth3
##      0.000000e+00      0.000000e+00      -6.372021e+01       0.000000e+00
##           ormoth4           ormoth5            orfath1            orfath2
##      0.000000e+00      0.000000e+00       0.000000e+00       0.000000e+00
##           orfath3           orfath4            orfath5             disllb
##      0.000000e+00      0.000000e+00       0.000000e+00       0.000000e+00
##           dtotord             dmar2          adequacy2          adequacy3
##      0.000000e+00     -4.846922e+01       0.000000e+00       0.000000e+00
##          nprevist        I(dmage^2)         I(dfage^2)        I(dmeduc^2)
##      0.000000e+00      0.000000e+00       0.000000e+00       0.000000e+00
##       I(dfeduc^2)       I(disllb^2)       I(dtotord^2)      I(nprevist^2)
##      0.000000e+00      0.000000e+00       0.000000e+00       0.000000e+00
##    tobacco2:csex2   tobacco2:mrace32   tobacco2:mrace33   tobacco2:preterm2
##      0.000000e+00      0.000000e+00       0.000000e+00       1.212800e+02
##    tobacco2:dmage     tobacco2:dfage    tobacco2:dmeduc     tobacco2:dfeduc
##      1.139206e+00      0.000000e+00       0.000000e+00       0.000000e+00
##  tobacco2:ormoth1   tobacco2:ormoth2   tobacco2:ormoth3    tobacco2:ormoth4
##      0.000000e+00      0.000000e+00       0.000000e+00       0.000000e+00
##  tobacco2:ormoth5   tobacco2:orfath1   tobacco2:orfath2    tobacco2:orfath3
##      0.000000e+00      0.000000e+00      -2.533540e+00       0.000000e+00
##  tobacco2:orfath4   tobacco2:orfath5    tobacco2:disllb    tobacco2:dtotord
##      0.000000e+00      0.000000e+00       0.000000e+00       9.383169e+00
##    tobacco2:dmar2 tobacco2:adequacy2 tobacco2:adequacy3   tobacco2:nprevist
##      0.000000e+00      1.233125e-01       0.000000e+00       1.262811e+00
##     csex2:mrace32      csex2:mrace33      csex2:preterm2         csex2:dmage
##      0.000000e+00      0.000000e+00       0.000000e+00       0.000000e+00
##       csex2:dfage       csex2:dmeduc       csex2:dfeduc       csex2:ormoth1
```

```
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##        csex2:ormoth2       csex2:ormoth3       csex2:ormoth4       csex2:ormoth5
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##        csex2:orfath1       csex2:orfath2       csex2:orfath3       csex2:orfath4
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##        csex2:orfath5        csex2:disllb       csex2:dtotord         csex2:dmar2
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##    csex2:adequacy2     csex2:adequacy3      csex2:nprevist     mrace32:preterm2
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##     mrace33:preterm2       mrace32:dmage       mrace33:dmage       mrace32:dfage
##        0.000000e+00        0.000000e+00        0.000000e+00       -7.236427e-01
##        mrace33:dfage      mrace32:dmeduc      mrace33:dmeduc      mrace32:dfeduc
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##      mrace33:dfeduc     mrace32:ormoth1     mrace33:ormoth1     mrace32:ormoth2
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##     mrace33:ormoth2     mrace32:ormoth3     mrace33:ormoth3     mrace32:ormoth4
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##     mrace33:ormoth4     mrace32:ormoth5     mrace33:ormoth5     mrace32:orfath1
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##     mrace33:orfath1     mrace32:orfath2     mrace33:orfath2     mrace32:orfath3
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##     mrace33:orfath3     mrace32:orfath4     mrace33:orfath4     mrace32:orfath5
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##     mrace33:orfath5      mrace32:disllb      mrace33:disllb     mrace32:dtotord
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##     mrace33:dtotord       mrace32:dmar2       mrace33:dmar2   mrace32:adequacy2
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##   mrace33:adequacy2   mrace32:adequacy3   mrace33:adequacy3    mrace32:nprevist
##        0.000000e+00        0.000000e+00        0.000000e+00       -8.111708e-02
##    mrace33:nprevist      preterm2:dmage      preterm2:dfage     preterm2:dmeduc
##        0.000000e+00        0.000000e+00        0.000000e+00        3.002415e+00
##     preterm2:dfeduc    preterm2:ormoth1    preterm2:ormoth2    preterm2:ormoth3
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##    preterm2:ormoth4    preterm2:ormoth5    preterm2:orfath1    preterm2:orfath2
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##    preterm2:orfath3    preterm2:orfath4    preterm2:orfath5     preterm2:disllb
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##    preterm2:dtotord      preterm2:dmar2 preterm2:adequacy2 preterm2:adequacy3
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##   preterm2:nprevist         dmage:dfage        dmage:dmeduc        dmage:dfeduc
##        1.984857e+01        0.000000e+00        0.000000e+00        0.000000e+00
##        dmage:ormoth1       dmage:ormoth2       dmage:ormoth3       dmage:ormoth4
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##        dmage:ormoth5       dmage:orfath1       dmage:orfath2       dmage:orfath3
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##        dmage:orfath4       dmage:orfath5        dmage:disllb       dmage:dtotord
##        0.000000e+00        0.000000e+00       -3.837685e-03        0.000000e+00
##          dmage:dmar2     dmage:adequacy2     dmage:adequacy3      dmage:nprevist
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##         dfage:dmeduc        dfage:dfeduc       dfage:ormoth1       dfage:ormoth2
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##        dfage:ormoth3       dfage:ormoth4       dfage:ormoth5       dfage:orfath1
##        0.000000e+00        0.000000e+00        0.000000e+00        0.000000e+00
##        dfage:orfath2       dfage:orfath3       dfage:orfath4       dfage:orfath5
```

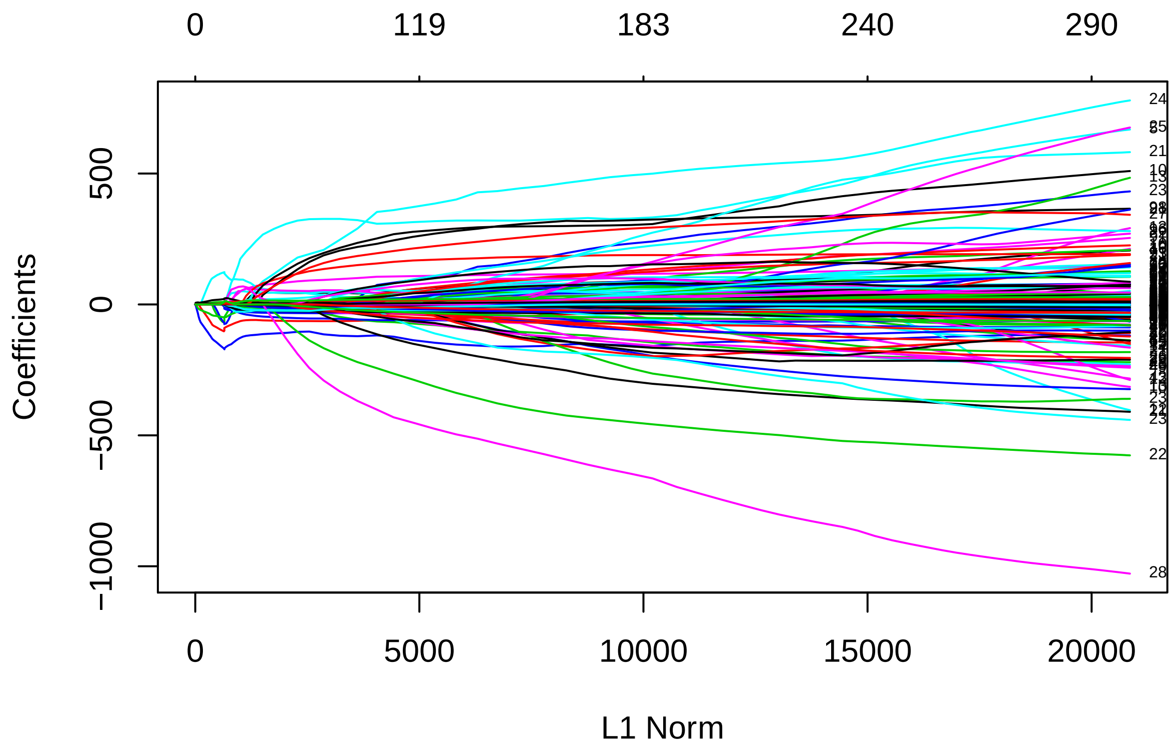
```
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##         dfage:disllb         dfage:dtotord           dfage:dmar2     dfage:adequacy2
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##      dfage:adequacy3       dfage:nprevist         dmeduc:dfeduc       dmeduc:ormoth1
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##       dmeduc:ormoth2       dmeduc:ormoth3        dmeduc:ormoth4       dmeduc:ormoth5
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##       dmeduc:orfath1       dmeduc:orfath2        dmeduc:orfath3       dmeduc:orfath4
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##       dmeduc:orfath5        dmeduc:disllb        dmeduc:dtotord          dmeduc:dmar2
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##     dmeduc:adequacy2     dmeduc:adequacy3       dmeduc:nprevist       dfeduc:ormoth1
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##       dfeduc:ormoth2       dfeduc:ormoth3        dfeduc:ormoth4       dfeduc:ormoth5
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##       dfeduc:orfath1       dfeduc:orfath2        dfeduc:orfath3       dfeduc:orfath4
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##       dfeduc:orfath5        dfeduc:disllb        dfeduc:dtotord         dfeduc:dmar2
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##     dfeduc:adequacy2     dfeduc:adequacy3       dfeduc:nprevist      ormoth1:orfath1
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##      ormoth2:orfath1      ormoth3:orfath1       ormoth4:orfath1      ormoth5:orfath1
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##      ormoth1:orfath2      ormoth2:orfath2       ormoth3:orfath2      ormoth4:orfath2
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##      ormoth5:orfath2      ormoth1:orfath3       ormoth2:orfath3      ormoth3:orfath3
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##      ormoth4:orfath3      ormoth5:orfath3       ormoth1:orfath4      ormoth2:orfath4
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##      ormoth3:orfath4      ormoth4:orfath4       ormoth5:orfath4      ormoth1:orfath5
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##      ormoth2:orfath5      ormoth3:orfath5       ormoth4:orfath5      ormoth5:orfath5
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##       ormoth1:disllb       ormoth2:disllb        ormoth3:disllb       ormoth4:disllb
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##       ormoth5:disllb      ormoth1:dtotord       ormoth2:dtotord      ormoth3:dtotord
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##      ormoth4:dtotord      ormoth5:dtotord         ormoth1:dmar2        ormoth2:dmar2
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##        ormoth3:dmar2        ormoth4:dmar2         ormoth5:dmar2    ormoth1:adequacy2
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##    ormoth2:adequacy2    ormoth3:adequacy2     ormoth4:adequacy2    ormoth5:adequacy2
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##    ormoth1:adequacy3    ormoth2:adequacy3     ormoth3:adequacy3    ormoth4:adequacy3
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##    ormoth5:adequacy3     ormoth1:nprevist      ormoth2:nprevist     ormoth3:nprevist
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##     ormoth4:nprevist     ormoth5:nprevist        orfath1:disllb       orfath2:disllb
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##       orfath3:disllb       orfath4:disllb        orfath5:disllb      orfath1:dtotord
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##      orfath2:dtotord      orfath3:dtotord       orfath4:dtotord      orfath5:dtotord
##         0.000000e+00         0.000000e+00         0.000000e+00         0.000000e+00
##        orfath1:dmar2        orfath2:dmar2         orfath3:dmar2        orfath4:dmar2
```

```
##       0.000000e+00       0.000000e+00       0.000000e+00       0.000000e+00
##      orfath5:dmar2  orfath1:adequacy2  orfath2:adequacy2  orfath3:adequacy2
##       0.000000e+00       0.000000e+00       0.000000e+00       0.000000e+00
##  orfath4:adequacy2  orfath5:adequacy2  orfath1:adequacy3  orfath2:adequacy3
##       0.000000e+00       0.000000e+00       0.000000e+00       0.000000e+00
##  orfath3:adequacy3  orfath4:adequacy3  orfath5:adequacy3   orfath1:nprevist
##       0.000000e+00       0.000000e+00       0.000000e+00       0.000000e+00
##   orfath2:nprevist   orfath3:nprevist   orfath4:nprevist   orfath5:nprevist
##       0.000000e+00       0.000000e+00       0.000000e+00       0.000000e+00
##     disllb:dtotord       disllb:dmar2   disllb:adequacy2   disllb:adequacy3
##      -2.014876e-02       0.000000e+00       0.000000e+00       0.000000e+00
##    disllb:nprevist      dtotord:dmar2  dtotord:adequacy2  dtotord:adequacy3
##       0.000000e+00       0.000000e+00       1.430495e+00       0.000000e+00
##   dtotord:nprevist    dmar2:adequacy2    dmar2:adequacy3     dmar2:nprevist
##       0.000000e+00       0.000000e+00       0.000000e+00       0.000000e+00
## adequacy2:nprevist adequacy3:nprevist
##       2.450708e+00       5.320214e+00
```

```
# print only non-zero lasso coefficients
print(reg_1c$beta[,25][reg_1c$beta[,25] != 0])
```

```
##              csex2            mrace32            mrace33            ormoth2
##      -1.004847e+02      -6.707636e+01      -1.645689e+02      -6.372021e+01
##              dmar2   tobacco2:preterm2     tobacco2:dmage    tobacco2:orfath2
##      -4.846922e+01       1.212800e+02       1.139206e+00      -2.533540e+00
##   tobacco2:dtotord tobacco2:adequacy2  tobacco2:nprevist      mrace32:dfage
##       9.383169e+00       1.233125e-01       1.262811e+00      -7.236427e-01
##   mrace32:nprevist    preterm2:dmeduc  preterm2:nprevist       dmage:disllb
##      -8.111708e-02       3.002415e+00       1.984857e+01      -3.837685e-03
##     disllb:dtotord  dtotord:adequacy2 adequacy2:nprevist adequacy3:nprevist
##      -2.014876e-02       1.430495e+00       2.450708e+00       5.320214e+00
```

```
# Plot what the coefficients are doing as we increase lambda
plot(reg_1c, label=TRUE)
```




```
# Each curve corresponds to a variable. It shows the path of its coefficient
#   against the L1-norm of the whole coefficient vector as lambda varies. The top axis
#   indicates the number of nonzero coefficients at the current lambda, which is the
#   effective degrees of freedom (df) for the lasso.
```

```
knitr::kable(reg_1c$beta[,25][reg_1c$beta[,25] != 0], caption=paste0("Lasso Regression for top (lambda =
             col.names = 'Non-zero Coefficients', align = "l", digits = 3)
```

Table 1: Lasso Regression for top (lambda = 12.2164603064059)

	Non-zero Coefficients
csex2	-100.485
mrace32	-67.076
mrace33	-164.569
ormoth2	-63.720
dmar2	-48.469
tobacco2:preterm2	121.280
tobacco2:dmage	1.139
tobacco2:orfath2	-2.534
tobacco2:dtotord	9.383
tobacco2:adequacy2	0.123
tobacco2:nprevist	1.263
mrace32:dfage	-0.724
mrace32:nprevist	-0.081
preterm2:dmeduc	3.002
preterm2:nprevist	19.849

	Non-zero Coefficients
dmage:disllb	-0.004
disllb:dtotord	-0.020
dtotord:adequacy2	1.430
adequacy2:nprevist	2.451
adequacy3:nprevist	5.320

Problem 2

Describe the propensity score approach to the problem of estimating the average causal effect of smoking when the treatment is randomly assigned conditional on the observables. How does it reduce the dimensionality problem of multivariate matching? Try a few ways to estimate the effects of maternal smoking on birthweight:

Part (a)

First create the propensity score. For our purposes let's use a logit specification. First specify the logit using all of the "predetermined" covariates (don't include interactions). Next, include only those "predetermined" covariates that enter significantly in the first logit specification. How comparable are the propensity scores? If they are similar does this imply that we have the "correct" set of covariates in the logit specification used for our propensity score?

```
# create the propensity score using logit
# using all of the "predetermined" covariates


# then try logit with only the significant covariates


# Compare histograms of p-scores
```

Part (b)

Control directly for the estimated propensity scores using a regression analysis, and estimate an average treatment effect. State clearly the assumptions under which your estimate is correct.

```
# Control for p-score in regression analysis

# Estimate ATE
```

Part (c)

As discussed in class, one can use the estimated propensity scores to reweight the outcomes of non- smokers and estimate the average treatment effect. Compute an estimate of the average treatment effect and the "effect of the treatment on the treated" by appropriate reweighting of the data.

```
# Reweight data using p-score to weight


# Estimate ATE
```

```
# Estimate TOT with rewieghted data
```

Part (d)

Estimate the counterfactual densities relevant for the above part with a kernel density estimator. That is, estimate the density of birthweight (or log birthweight) if everyone smoked and again if no one smoked. Hint: Consider directly applying the Hirano, Imbens, and Ridder propensity score reweighting scheme in the context of estimating the densities of the treated and control groups (rather than the means of the treated and control groups). Stata has very useful preprogrammed commands. In addition to using the preprogrammed Stata command to compute/graph the kernel density over the entire range of birthweight, please also calculate by hand the kernel estimator at birthweight equals 3,000 grams (and provide the code you wrote that shows the calculation of the kernel estimator at this single point). Play around with a bandwidth starting with half the default Stata bandwidth. Choose the same bandwidth for all the pictures, and produce a (beautiful, production quality) figure depicting both densities.

```
# Estimate the counterfactual birthweight densities with a kernel density estimator
# See Joel's notes for kernel density estimator
# Play around with a bandwidth starting with half the default Stata bandwidth
# For stata bandwidht, see rkdensity.pdf page 9 in this ps1b github folder.
# You can also run on stata with no bandwidth specified, then print the
# default bandwidth used using `display r(bwidth)`
# Choose the same bandwidth for all the pictures


# Graph both kernel densities over range of birthweight in the same plot


# calculate the kernel estimator at birthweight equals 3,000 grams
```

Part (e)

Take one of your densities and display an estimate of the density using different bandwidths as well as the one you settled on. What happens with bigger (smaller) bandwidths?

Part (f)

What are the benefits of the weighting approach (from part c)? What are the potential drawbacks? Pay particular attention to to the issue of people with extremely high and extremely low values of the propensity score.

Part (g)

Present your findings and interpret the results on the relationship between birthweight and smoking. For the estimates in parts (b) and (c), consider which of the following conditions must hold in order for that estimate to be valid:

- The treatment effect heterogeneity is linear in the propensity score.
- The treatment effect heterogeneity is not linear in the propensity score.
- The decision to smoke is completely randomly assigned.
- Conditional on the exogenous variables the decision to smoke is randomly assigned.

Problem 3

A potentially more informative way to describe how birth weight affects smoking is to estimate the "non-parametric" conditional mean of birth weight as a function of the estimated probability of smoking, separately for smokers and non-smokers on the same graph. To do so, divide the data from smokers into 100 approximately equally spaced bins based on the estimated propensity score. Do the same for nonsmokers. Use the blocking estimator we discussed in class. Interpret your findings and relate them to the results in (2b).

Problem 4

Low birth weight births (less than 2500 grams) are considered particularly undesirable since they comprise a large share of infant deaths. Redo question 3 using an indicator for low birth weight birth as the outcome of interest. Interpret your findings.

Problem 5

Let's link matching back to regression. Consider the conditional expectation function $\mathbb{E}[birthweight \mid X]$, where X contains the following variables: `rectype pldel3 cntocpop stresfip dmage mrace3 dmar adequacy csex dplural`.

Part (a)

Develop a regression that you are confident estimates $\mathbb{E}[birthweight \mid X]$ as $N \to \infty$? Why are you confident that your regression gets the CEF right?

```
# Select variables
df5a = df %>% select(dbrwt, rectype, pldel3, cntocpop, stresfip,
                     dmage, mrace3, dmar, adequacy, csex, dplural)

# Run regression
```

Part (b)

Now run the regression you propose above, but add the treatment (your binary smoking variable) as the righthand side variable of interest. Prove that if the treatment effect of smoking on birthweight is independent of the covariates in X, then exact matching and your regression estimate the same thing. You may assume the conditional independence assumption holds given the variables in X listed above.

```
# Select vars and smoking indicator
df5b = df %>% select(dbrwt, rectype, pldel3, cntocpop, stresfip,
                     dmage, mrace3, dmar, adequacy, csex, dplural)

# Run regression
```

Part (c)

Develop a weighted version of the exact matching estimator that estimates the same thing as the regression above (regardless of whether the treatment effect is independent of covariates).

Part (d)

Estimate the weighted matching estimator you propose. Compare it to the regression estimate from part (b). Are they similar?

Part (e)

Is the sample size of your regression the same as the sample size of your matching estimator, or does the regression have more observations? If the regression has more observations, why don't these extra observations influence the treatment effect estimate?

Part (f)

Compute a standard error for your matching estimator using the formula from Imbens (2015). Specifically, note that your matching estimator should have a form

$$\frac{1}{N_t}\sum_{d_i=1} w_i y_i - \frac{1}{N_c}\sum_{d_i=0} w_i y_i$$

where $\sum_{d_i=1} w_i = N_t$ and $\sum_{d_i=0} w_i = N_c$. Then the conditional variance is approximately

$$\sum_i \left(\frac{d_i}{N_t^2} + \frac{1-d_i}{N_c^2} w_i^2 \hat{\sigma}^2_{d_i}(x_i) \right),$$

where $\hat{\sigma}^2_{d_i}(x_i) = \frac{1}{2}(y_i - y_{nn(i)})$, and $y_{nn(i)}$ is the nearest neighbor to observation i with the *same* treatment status. Figure out the implicit weights w_i in your estimator from part (d), and compute the conditional variance. Is it close to your regression coefficient variance?

```
# Compute a standard error for your matching estimator using the formula from Imbens (2015).


# compute the conditional variance of estimator from (d)
```

Problem 6

Concisely and coherently summarize your overall results, providing some intuition. Write it like you would the conclusion of a paper. In this summary, describe whether you think your best estimate of the effects of smoking is credibly identified. State why or why not.