Seminar 10: Machine Learning for Microeconometrics

A Monte Carlo exercise: OLS vs Naive Lasso vs Double Selection LASSO

We want to check how the different methods work.

Setup of the Monte Carlo Simulation

The following simulation is a slightly revised version of the simulation from S. Kranz "Lasso and the Methods of Causality" (https://skranz.github.io/r/2020/09/14/LassoCausality.html)

We study different approaches for variable selection using a Monte-Carlo simulation with the following data generating process for

$$y = 1 + \alpha d + \sum_{k=1}^{K_c} \beta_k^{cy} x_k^c + \sum_{k=1}^{K_y} \beta_k^y x_k^y + \varepsilon^y$$

$$d = 1 + \sum_{k=1}^{K_c} \beta_k^{cd} x_k^c + \sum_{k=1}^{K_e} \beta_k^e x_k^e + \varepsilon^d$$

We have the following potential control variables that are all independently normally distributed from each other:

- d the endogenous variable
- x_k^c is one of K_c confounders that directly affect both d and y. For a consistent OLS estimation of α, we need to control for all confounders.
- x_k^y is one of K_y variables that only affect the dependent variable y but not the explanatory variable d. Whether we would add it or not in an OLS regression should not affect the bias of our estimator $\hat{\alpha}$.
- x_k^e is one of K_e variables that only affect d but not through any other channel the dependent variable y. It constitutes a source of exogenous variation in d. We can estimate in an OLS regression α more precisely if we don't add any x_k^e to the regression. Also, we will see that adding fewer x_k^e can reduce the bias of an OLS estimator that arises if we have not perfectly controlled for all confounders.
- We also observe K_u variables x_k^u that neither affect y nor d. They are just uncorrelated noise variables that we ideally leave out of our regressions.

The following code simulates a data set with $n = 200$ observations, $K_c = 15$ confounders, $K_y = 15$ variables that affect only y, $K_e = 2$ that provides a source of exogenous variation and $K_u = 200$ explanatory variables that are uncorrelated with everything else. The causal effect of interest α, all other regression coefficients and standard deviations are equal to 1.

In [ ]:
```
import stata_setup
stata_setup.config("C:/Program Files/Stata17", "se")
```

In [ ]:
```
%%stata

clear all
set seed 20211109
set linesize 255

set obs 200

forvalues k=1/232{
                g V`k'=rnormal(0,1)
        }

egen sum_c=rowtotal(V1-V15)
egen sum_y=rowtotal(V16-V30)
egen sum_e=rowtotal(V31-V32)

g d=1+1*sum_c+sum_e+rnormal()
g y=1+d+sum_c+sum_y+rnormal()
```

OLS

We look at five different cases.

- kitchen sink specification (throw in all potential variables)
- Omitted Variable Bias
- Data Generating Process
- controlling variables that correlated with D but not impact y via other channels (valid IV for D)
- IV to correct OMB

In [4]:
```
%%stata
quietly reg y d V*
estimates store r1, title(kitchen)

qui{
quietly reg y d
estimates store r2, title(OVB)

quietly reg y d V1-V30
estimates store r3, title(DGP)

quietly reg y d V1-V32
estimate store r4, title(D_e)

quietly ivregress 2sls y V16-V30 (d=V31-V32)
estimates store r5, title(IV)
}

estout r*, keep(d) cells(b se(par fmt(2))) ///
   legend label varlabels(_cons Constant)
```

```
. quietly reg y d V*

. estimates store r1, title(kitchen)

.
. qui{

.
. estout r*, keep(d) cells(b se(par fmt(2))) ///
>    legend label varlabels(_cons Constant)

------------------------------------------------------------------------------------
                          kitchen          OVB           DGP           D_e           IV
                             b/se         b/se          b/se          b/se         b/se
------------------------------------------------------------------------------------
d                        1.987665     1.694304      .9467929      .9702836     .5892811
                              (.)       (0.07)        (0.05)        (0.07)       (0.44)
------------------------------------------------------------------------------------

.
.
```

LASSO, Post-Lasso and Post Double Selection

In the following we estimate the relationship between y and predictors (D and V^*). The goal is to recover the causal parameter α when there are high numbers of potential controls. (NOTE: we assume thta CIA holds.)

As we discussed in the lecture, the lasso regression selects a subset of explanatory variables whose estimated coefficients are non-zero. But also the coefficients of the selected variables will be typically attenuated towards 0 because of the penalty term. The post-lasso estimator avoids this attenuation by simply performing an OLS estimation using all the selected variables from the lasso estimation.

In the following, we use the stata package "lassopack" and "pdslasso".

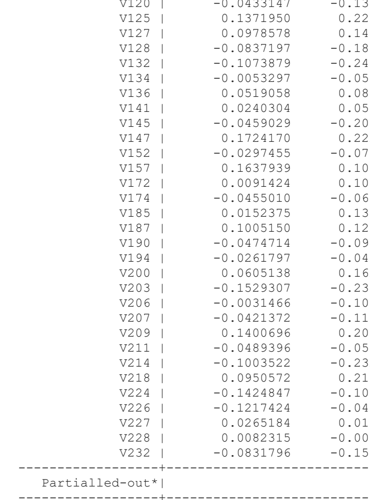
LASSO and Post-Lasso

We use three different procedures to determine the penalty parameter λ.

1. cross validation (cvlasso)
2. using information criterion (lasso2)
3. Theoretical value (rlasso)

In [5]:
```
%%stata

cvlasso y d V*, lopt  postres plotcv
```

```
.
. cvlasso y d V*, lopt  postres plotcv

K-fold cross-validation with 10 folds. Elastic net with alpha=1.
Fold 1 2 3 4 5 6 7 8 9 10
          |          Lambda           MSPE       st. dev.
----------+---------------------------------------------
         1|       2875.711      70.645657      11.235848
         2|      2620.2408      64.216646      11.392625
         3|      2387.4658      56.658243      10.001639
         4|      2175.3699      50.379822      8.8255175
         5|       1982.116      45.164359      7.8306878
         6|      1806.0303      40.831648      6.9890582
         7|      1645.5876      37.232059      6.2770763
         8|      1499.3981      34.241341      5.6749487
         9|      1366.1957      31.756322      5.1659934
        10|      1244.8266      29.691325      4.7361058
        11|      1134.2397      27.975208      4.3733211
        12|       1033.477      26.548889      4.0674591
        13|      941.66571      25.363304      3.8098367
        14|      858.01072      24.377709      3.5930362
        15|      781.78741      23.558264       3.410719
        16|      712.33557      22.876866      3.2574725
        17|      649.05364      22.310172       3.128684
        18|      591.39351      21.838848      3.0204264
        19|      538.85574      21.449187      2.9289673
        20|      490.98528      20.978068      2.8658107
        21|       447.3675      20.001976       2.751707
        22|       407.6246      18.857646      2.6025543
        23|      371.41236      17.795525      2.4442804
        24|      338.41711       16.72517      2.2863742
        25|      308.35307      15.607381      2.1197607
        26|      280.95983      14.508302      1.9443237
        27|      256.00013      13.347922      1.7532399
        28|      233.25779      12.165617      1.5734637
        29|      212.53581      10.939476      1.4209356
        30|      193.65471      9.7231545      1.2727875
        31|      176.45096      8.6372147      1.1059851
        32|      160.77554      7.6728737      .95771969
        33|      146.49269      6.8423957      .82191562
        34|      133.47868       6.154876      .70758563
        35|       121.6208       5.576866      .61325231
        36|      110.81635      5.0997055      .53452071
        37|      100.97173      4.7059639      .47054332
        38|      92.001677      4.3705292      .41351908
        39|      83.828502      4.0886482      .36621063
        40|      76.381409      3.8508632      .32496497
        41|      69.595895      3.6448411      .29419368
        42|      63.413187      3.4825931      .27209035
        43|      57.779735      3.3321031      .25195832
        44|      52.646742      3.1990837      .23413112
        45|      47.969751      3.0825292      .22083539
        46|       43.70825      2.9914454      .21557341
        47|       39.82533      2.9094441      .20837932
        48|      36.287357      2.8407593      .20120308  ^
        49|      33.063688      2.7765465      .19469713
        50|      30.126401      2.7273386      .18686482
        51|      27.450055      2.6888654      .18303298
        52|      25.011468      2.6734949      .18286842  *
        53|      22.789518      2.6801621      .18520002
        54|       20.76496      2.6854249      .18826565
        55|      18.920258      2.7049528      .19174921
        56|      17.239435      2.7429811      .19776096
        57|      15.707931      2.8041509      .20779702
        58|      14.312481      2.8846656      .22185323
        59|         13.041      2.9667291      .23929786
        60|      11.882474       3.062136       .2561358
        61|      10.826867       3.158131      .27549286
        62|      9.8650383      3.2604377       .2975459
        63|      8.9886555      3.3429892      .31992363
        64|      8.1901281      3.4294875      .33957761
        65|      7.4625397      3.5160464      .35482796
        66|      6.7995882      3.6165174      .36713824
        67|      6.1955315      3.7184379      .38266566
        68|      5.6451376      3.8252545      .40195705
        69|      5.1436391      3.9400959      .42233998
        70|      4.6866924      4.0623193      .44629588
        71|      4.2703396      4.1763518      .46576863
        72|      3.8909744      4.2896207      .48256248
        73|      3.5453109      4.4222282      .50055401
        74|      3.2303553      4.5759979      .51888628
        75|      2.9433794      4.7444922      .53194604
        76|      2.6818977      4.9214113      .53686173
        77|      2.4436453      5.0942629      .54161414
        78|      2.2265586       5.254475      .54620122
* lopt = the lambda that minimizes MSPE.
  Run model: cvlasso, lopt
^ lse = largest lambda for which MSPE is within one standard error of the minimal MSPE.
  Run model: cvlasso, lse
Estimate lasso with lambda=25.011 (lopt).

---------------------------------------------------
         Selected |           Lasso   Post-est OLS
------------------+--------------------------------
                d |       1.7671647      1.6330105
               V2 |       0.2170179      0.4018278
               V3 |       0.1438550      0.2795591
               V4 |       0.1615942      0.2587519
               V5 |       0.1262491      0.3055922
               V7 |       0.1710275      0.4392129
               V8 |       0.2960553      0.4429216
               V9 |       0.2431908      0.4571931
              V10 |       0.1001006      0.2621215
              V11 |       0.1318062      0.3447644
              V12 |       0.2067372      0.3165472
              V13 |       0.1509648      0.3731960
              V14 |       0.1483965      0.4051502
              V15 |       0.0047646      0.2698030
              V16 |       0.9566156      1.0253114
              V17 |       0.8971065      0.9708301
              V18 |       0.8662542      0.8547626
              V19 |       0.8416409      0.8938764
              V20 |       0.9733480      1.0402765
              V21 |       0.9984264      1.1270631
              V22 |       0.8746915      0.9466926
              V23 |       0.7131791      0.7185309
              V24 |       0.8926904      0.9730758
              V25 |       0.9660915      0.9665033
              V26 |       0.9687728      1.0361031
              V27 |       1.0507142      1.2212056
              V28 |       0.9852362      0.9878084
              V29 |       0.8618398      0.9939550
              V30 |       0.9288164      1.0054000
              V31 |      -0.4784266     -0.4581066
              V32 |      -0.6452484     -0.5223352
              V34 |      -0.0363806     -0.1394978
              V38 |      -0.0084473     -0.0938583
              V45 |       0.0015656      0.0977555
              V52 |      -0.0648039     -0.1967482
              V54 |      -0.1287554     -0.1463557
              V56 |      -0.0452706     -0.1312493
              V57 |      -0.0057927     -0.0307171
              V60 |      -0.0422042     -0.0490789
              V61 |       0.0052681      0.1021010
              V66 |       0.1242700      0.0991859
              V73 |       0.0045445      0.1003533
              V76 |       0.0724398      0.0652381
              V81 |       0.0554979      0.0213331
              V88 |       0.1363966      0.2787004
              V90 |      -0.0167349     -0.0216251
              V92 |      -0.0470425     -0.1255307
              V95 |      -0.1486033     -0.1431532
              V99 |      -0.0679617     -0.1253309
             V100 |      -0.1897291     -0.2020689
             V103 |       0.0303305      0.1251649
             V106 |       0.1611949      0.2391026
             V107 |      -0.0452993     -0.0413112
             V112 |      -0.0172567     -0.0772297
             V117 |      -0.1276827     -0.2163777
             V120 |      -0.0433147     -0.1350839
             V125 |       0.1371950      0.2268413
             V127 |       0.0978578      0.1466474
             V128 |      -0.0837197     -0.1895606
             V132 |      -0.1073879     -0.2474157
             V134 |      -0.0053297     -0.0552603
             V136 |       0.0519058      0.0840987
             V141 |       0.0240304      0.0589167
             V145 |      -0.0459029     -0.2041398
             V147 |       0.1724170      0.2223354
             V152 |      -0.0297455     -0.0734237
             V157 |       0.1637939      0.1047062
             V172 |       0.0091424      0.1006885
             V174 |      -0.0455010     -0.0615754
             V185 |       0.0152375      0.1344090
             V187 |       0.1005150      0.1293806
             V190 |      -0.0474714     -0.0963638
             V194 |      -0.0261797     -0.0418876
             V200 |       0.0605138      0.1628385
             V203 |      -0.1529307     -0.2376098
             V206 |      -0.0031466     -0.1019382
             V207 |      -0.0421372     -0.1153863
             V209 |       0.1400696      0.2085899
             V211 |      -0.0489396     -0.0530495
             V214 |      -0.1003522     -0.2329026
             V218 |       0.0950572      0.2134265
             V224 |      -0.1424847     -0.1032286
             V226 |      -0.1217424     -0.0495117
             V227 |       0.0265184      0.0130600
             V228 |       0.0082315     -0.0034970
             V232 |      -0.0831796     -0.1557677
------------------+--------------------------------
  Partialled-out* |
------------------+--------------------------------
            _cons |       0.3432482      0.4837121
---------------------------------------------------

.
```



In [6]:
```
%%stata

lasso2 y d V*, lic(aicc) postres
```

```
.
. lasso2 y d V*, lic(aicc) postres

  Knot|  ID     Lambda    s      L1-Norm        AICc     R-sq   | Action
------+---------------------------------------------------------+---------------
     1|   1 2875.71100     0     0.00000    854.03000   0.0000  | Added _cons.
     2|   2 2620.24076     1     0.15052    829.60182   0.1239  | Added d.
     3|  20  490.98528     5     1.63202    611.78416   0.7172  | Added V16 V17 V21 V24.
     4|  21  447.36750     7     2.14861    601.38752   0.7372  | Added V20 V22.
     5|  22  407.62460     8     2.74113    587.37607   0.7577  | Added V23.
     6|  23  371.41236    10     3.36033    575.15009   0.7770  | Added V27 V28.
     7|  24  338.41711    12     4.05026    561.61482   0.7963  | Added V18 V25.
     8|  25  308.35307    13     4.83485    543.21340   0.8163  | Added V19.
     9|  26  280.95983    15     5.56418    528.54364   0.8333  | Added V26 V107.
    10|  27  256.00013    16     6.36685    509.65979   0.8501  | Added V29.
    11|  28  233.25779    22     7.26676    500.43341   0.8671  | Added V8 V11 V30 V32 V174 V232.
    12|  29  212.53581    24     8.28284    477.49842   0.8845  | Added V12 V224.
    13|  30  193.65471    25     9.26824    451.49573   0.8999  | Added V31.
    14|  31  176.45096    26    10.17767    426.39476   0.9129  | Added V100.
    15|  34  133.47868    27    12.46738    353.91247   0.9402  | Added V157.
    16|  35  121.62080    28    13.12965    333.79601   0.9466  | Added V226.
    17|  36  110.81635    30    13.74641    317.83076   0.9521  | Added V4 V209.
    18|  37  100.97173    33    14.32838    306.01199   0.9567  | Added V9 V66 V187.
    19|  39   83.82850    38    15.40839    283.18608   0.9642  | Added V54 V95 V106 V147 V203.
    20|  40   76.38141    42    15.95030    277.25318   0.9673  | Added V2 V13 V57 V141.
    21|  41   69.59589    46    16.48793    272.47368   0.9702  | Added V10 V81 V125 V127.
    22|  42   63.41319    49    17.01652    264.92454   0.9727  | Added V76 V166 V218.
    23|  43   57.77973    52    17.54675    257.80016   0.9751  | Added V117 V211 V228.
    24|  44   52.64674    56    18.07165    255.32185   0.9772  | Added V5 V60 V88 V132.
    25|  45   47.96975    58    18.62455    244.99778   0.9792  | Added V3 V7.
    26|  46   43.70825    62    19.19593    243.13349   0.9810  | Added V14 V99 V152 V214.
    27|  47   39.82533    68    19.77635    251.45290   0.9827  | Added V56 V92 V128 V190 V194 V227.
    28|  48   36.28736    74    20.36457    262.09153   0.9842  | Added V52 V103 V120 V136 V200 V207.
    29|  49   33.06369    75    20.95376    248.61946   0.9856  | Added V34.
    30|  50   30.12640    78    21.52386    246.53473   0.9869  | Added V90 V112 V145.
    31|  51   27.45006    79    22.05532    235.33670   0.9879  | Added V172 V185. Removed V166.
    32|  52   25.01147    86    22.58345    260.37084   0.9889  | Added V15 V38 V45 V61 V73 V134 V206.
    33|  53   22.78952    95    23.18821    303.50036   0.9898  | Added V33 V46 V86 V139 V142 V149 V150 V216 V2
25.
    34|  54   20.76496    98    23.78930    308.39030   0.9907  | Added V1 V116 V123.
    35|  55   18.92026   104    24.43193    339.59114   0.9915  | Added V44 V77 V113 V140 V168 V221 V223. Remov
ed V228.
    36|  56   17.23943   108    25.05128    358.35424   0.9923  | Added V39 V75 V94 V193.
    37|  57   15.70793   117    25.75127    435.01964   0.9930  | Added V6 V65 V79 V115 V126 V135 V153 V167 V17
7 V189. Removed V57.
    38|  58   14.31248   117    26.51031    414.37760   0.9937  | Added V68. Removed V189.
    39|  59   13.04100   121    27.22867    444.88868   0.9943  | Added V57 V163 V170 V229.
    40|  60   11.88247   123    27.91090    453.10342   0.9948  | Added V114 V162 V186. Removed V216.
    41|  61   10.82687   130    28.61847    540.31246   0.9953  | Added V84 V87 V93 V124 V133 V161 V189 V199. R
emoved V113.
    42|  62    9.86504   139    29.33753    693.94399   0.9957  | Added V70 V85 V97 V101 V111 V137 V148 V205 V2
08 V230. Removed V211.
    43|  63    8.98866   140    30.07366    695.33162   0.9962  | Added V49 V196 V204. Removed V87 V194.
    44|  64    8.19013   144    30.77759    773.36942   0.9965  | Added V36 V74 V89 V119 V171 V215. Removed V68
V79.
    45|  65    7.46254   146    31.47698    807.46395   0.9969  | Added V43 V58 V64 V67 V96 V211. Removed V120
V189 V206 V208.
    46|  66    6.79959   146    32.12481    787.77313   0.9972  | Added V51 V79 V182. Removed V38 V60 V70.
    47|  67    6.19553   150    32.76044    891.44107   0.9974  | Added V35 V62 V108 V154 V178. Removed V79.
    48|  68    5.64514   152    33.36608    942.12173   0.9977  | Added V40 V82 V151. Removed V168.
    49|  69    5.14364   155    33.95511   1039.71459   0.9979  | Added V59 V71 V102 V105 V202. Removed V45 V11
6.
    50|  70    4.68669   155    34.51296   1022.00988   0.9981  | Added V69 V72 V143. Removed V73 V81 V225.
    51|  71    4.27034   158    35.02845   1138.63030   0.9982  | Added V78 V91 V181.
    52|  72    3.89097   162    35.55565   1332.36251   0.9984  | Added V37 V79 V130 V164 V192. Removed V151.
    53|  73    3.54531   165    36.08005   1506.20897   0.9985  | Added V109 V180 V188 V222. Removed V230.
    54|  74    3.23036   170    36.65839   1894.03936   0.9986  | Added V104 V155 V184 V194 V216.
    55|  75    2.94338   169    37.30700   1782.66548   0.9987  | Added V173 V210. Removed V56 V62 V72.
    56|  76    2.68190   171    37.86633   1955.91045   0.9988  | Added V48 V56 V83 V122 V144. Removed V44 V133
V193.
    57|  77    2.44365   175    38.45121   2414.11982   0.9989  | Added V38 V42 V80 V146 V158 V195. Removed V66
V182.
    58|  78    2.22656   178    39.09824   2870.17419   0.9990  | Added V120 V131 V138 V179 V182 V193 V197 V22
0. Removed V43 V78 V95 V195 V227.
Use long option for full output.

Use lambda=27.4500550129593 (selected by AICC).

---------------------------------------------------
         Selected |           Lasso   Post-est OLS
------------------+--------------------------------
                d |       1.7739059      1.7199908
               V2 |       0.2037073      0.3511454
               V3 |       0.1329859      0.2579224
               V4 |       0.1572871      0.2049401
               V5 |       0.1146634      0.2349067
               V7 |       0.1506317      0.3518616
               V8 |       0.2870204      0.3663435
               V9 |       0.2266119      0.4079307
              V10 |       0.0910298      0.1606296
              V11 |       0.1202902      0.2304665
              V12 |       0.2020258      0.2575092
              V13 |       0.1367833      0.2675648
              V14 |       0.1262425      0.3547255
              V16 |       0.9503031      1.0150096
              V17 |       0.8935241      0.8881534
              V18 |       0.8653531      0.8987193
              V19 |       0.8359564      0.9145585
              V20 |       0.9689720      1.0161326
              V21 |       0.9858095      1.1178121
              V22 |       0.8661082      0.9654338
              V23 |       0.7123804      0.7247260
              V24 |       0.8865819      0.9591149
              V25 |       0.9637724      1.0070754
              V26 |       0.9594685      1.0811744
              V27 |       1.0341885      1.2132579
              V28 |       0.9845314      0.9961394
              V29 |       0.8496424      0.9834720
              V30 |       0.9169324      1.0655068
              V31 |      -0.4768019     -0.5119249
              V32 |      -0.6499044     -0.6250956
              V34 |      -0.0259141     -0.1306937
              V52 |      -0.0534657     -0.1940602
              V54 |      -0.1242312     -0.1943650
              V56 |      -0.0371161     -0.1088089
              V57 |      -0.0070057      0.0028170
              V60 |      -0.0401164     -0.0670297
              V66 |       0.1247185      0.1430068
              V76 |       0.0711598      0.1029364
              V81 |       0.0560424      0.0595773
              V88 |       0.1257860      0.2370584
              V90 |      -0.0141130     -0.0738091
              V92 |      -0.0423251     -0.1067082
              V95 |      -0.1463076     -0.1865202
              V99 |      -0.0615865     -0.1386444
             V100 |      -0.1872265     -0.2223115
             V103 |       0.0232429      0.0914264
             V106 |       0.1526630      0.2510607
             V107 |      -0.0469361     -0.0127425
             V112 |      -0.0099131     -0.0812417
             V117 |      -0.1190439     -0.2038928
             V120 |      -0.0358203     -0.1097716
             V125 |       0.1279413      0.2269295
             V127 |       0.0918071      0.1662350
             V128 |      -0.0712827     -0.2219974
             V132 |      -0.0960809     -0.2075333
             V136 |       0.0451453      0.1469081
             V141 |       0.0235848      0.0228998
             V145 |      -0.0315228     -0.1946050
             V147 |       0.1665993      0.2307583
             V152 |      -0.0275227     -0.0562443
             V157 |       0.1684175      0.1186611
             V172 |       0.0019873      0.0904462
             V174 |      -0.0449817     -0.0361272
             V185 |       0.0045770      0.1132032
             V187 |       0.0978920      0.1216176
             V190 |      -0.0424959     -0.1043682
             V194 |      -0.0232709     -0.0645991
             V200 |       0.0500716      0.1827080
             V203 |      -0.1429959     -0.2563450
             V207 |      -0.0332539     -0.1261842
             V209 |       0.1318814      0.2227550
             V211 |      -0.0485428     -0.0466908
             V214 |      -0.0878162     -0.2204309
             V218 |       0.0844119      0.1795192
             V224 |      -0.1463122     -0.0853862
             V226 |      -0.1249878     -0.0984599
             V227 |       0.0241029      0.0696838
             V228 |       0.0097119     -0.0153371
             V232 |      -0.0779676     -0.1196546
------------------+--------------------------------
  Partialled-out* |
------------------+--------------------------------
            _cons |       0.3373338      0.3820430
---------------------------------------------------

.
```

In [7]:
```
%%stata

rlasso y d V*
```

```
.
. rlasso y d V*

---------------------------------------------------
         Selected |           Lasso   Post-est OLS
------------------+--------------------------------
                d |       1.3984524      1.7373940
              V17 |       0.0878541      1.3772976
              V24 |       0.0252680      1.2994967
            _cons |*      0.6117565      0.2027828
---------------------------------------------------
*Not penalized

.
```

Double Selection Lasso

The post double selection method by Belloni et. al. (2014) selects the control variables as follows. We run two lasso regressions.

- The first regresses d on all potential controls.
- The second regresses y on all potential controls (excluding d).
- Then we use the union of the selected variables from both lasso regressions for our post-lasso OLS regression.

An intuition for this approach is that confounders are variables that affect both d and y. To ensure that very few confounders are omitted, it thus seems not implausible to use as a broad control set all variables that relevantly affect d or y.

Let us apply the post double selection method:

In [8]:
```
%%stata
    quietly rlasso y V*
        local string1 = e(selected)
        di "`string1'"

    quietly rlasso d V*
        local string2 = e(selected)
        di "`string2'"

        local newlist: list string1 | string2
        di "`newlist'"

         reg y d `newlist'
```

```
.     quietly rlasso y V*

.         local string1 = e(selected)

.         di "`string1'"
V2 V8 V9 V10 V11 V13 V14 V15

.
.     quietly rlasso d V*

.         local string2 = e(selected)

.         di "`string2'"
V2 V3 V6 V8 V9 V10 V11 V12 V13 V14 V15 V25

.
.         local newlist: list string1 | string2

.         di "`newlist'"
V2 V8 V9 V10 V11 V13 V14 V15 V3 V6 V12 V25

.
.          reg y d `newlist'

      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(13, 186)      =     47.98
       Model |  10910.1609        13  839.243148   Prob > F        =    0.0000
    Residual |  3253.24827       186  17.4905821   R-squared       =    0.7703
-------------+----------------------------------   Adj R-squared   =    0.7543
       Total |  14163.4092       199  71.1729105   Root MSE        =    4.1822

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           d |   1.443783   .1290168    11.19   0.000     1.189259    1.698308
          V2 |   .0304683   .3343658     0.09   0.927    -.6291686    .6901051
          V8 |   .6232437   .3090462     2.02   0.045     .0135573     1.23293
          V9 |   .8211188   .3271792     2.51   0.013     .1756596    1.466578
         V10 |   .5466191   .3139174     1.74   0.083    -.0726771    1.165915
         V11 |   .6675631   .3052216     2.19   0.030     .0654218    1.269704
         V13 |   .3324538   .3463546     0.96   0.338    -.3508345    1.015742
         V14 |   .4096044   .3084713     1.33   0.186    -.1989479    1.018157
         V15 |      .3454   .3436906     1.00   0.316    -.3326328    1.023433
          V3 |   .0200068   .3257529     0.06   0.951    -.6226387    .6626522
          V6 |   .1951426   .3233501     0.60   0.547    -.4427625    .8330477
         V12 |   .7539879   .3212112     2.35   0.020     .1203024    1.387673
         V25 |   1.149046   .3026321     3.80   0.000     .5520135    1.746079
       _cons |   .5076797   .3261692     1.56   0.121    -.1357869    1.151146
------------------------------------------------------------------------------

.
```

Another closed related alternative is to apply the orthogonal score functions, as we discussed in the lecture (the double machine learning method). In our case, it is equivalent to regress the residual (of y on the selected regressors) on the residual (of d on the selected regressors). We also apply this method:

In [9]:
```
%%stata
    capture drop u v
    quietly reg y `string1'
    predict u, res

    quietly reg d `string2'
    predict v, res

    reg u v
```

```
.     capture drop u v

.     quietly reg y `string1'

.     predict u, res

.
.     quietly reg d `string2'

.     predict v, res

.
.     reg u v

      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(1, 198)       =     90.13
       Model |  2190.35979         1  2190.35979   Prob > F        =    0.0000
    Residual |  4811.83943       198  24.3022194   R-squared       =    0.3128
-------------+----------------------------------   Adj R-squared   =    0.3093
       Total |  7002.19923       199  35.1869308   Root MSE        =    4.9297

------------------------------------------------------------------------------
           u | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           v |   1.443783   .1520783     9.49   0.000     1.143882    1.743684
       _cons |   4.76e-09   .3485844     0.00   1.000    -.6874146    .6874146
------------------------------------------------------------------------------

.
```

Or we can use the "pdslasso" command

In [10]:
```
%%stata

pdslasso y d (V*)
```

```
.
. pdslasso y d (V*)
1.  (PDS/CHS) Selecting HD controls for dep var y...
Selected: V2 V8 V9 V10 V11 V13 V14 V15
2.  (PDS/CHS) Selecting HD controls for exog regressor d...
Selected: V2 V3 V6 V8 V9 V10 V11 V12 V13 V14 V15 V25


Estimation results:

Specification:
Regularization method:                 lasso
Penalty loadings:                      homoskedastic
Number of observations:                200
Exogenous (1):                         d
High-dim controls (232):               V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20 V
21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36 V37 V38 V39 V40 V41 V42 V43 V44 V45 V46 V47 V48
V49 V50 V51 V52 V53 V54 V55 V56
                                       V57 V58 V59 V60 V61 V62 V63 V64 V65 V66 V67 V68 V69 V70 V71 V72 V73 V74
V75 V76 V77 V78 V79 V80 V81 V82 V83 V84 V85 V86 V87 V88 V89 V90 V91 V92 V93 V94 V95 V96 V97 V98 V99 V100 V101 V
102 V103 V104 V105 V106 V107                                            V57 V58 V59 V60 V61 V62 V63 V64 V65 V66 V67
V68 V69 V70 V71 V72 V73 V74 V75 V76 V77 V78 V79 V80 V81 V82 V83 V84 V85 V86 V87 V88 V89 V90 V91 V92 V93 V94 V95
V96 V97 V98 V99 V100 V101 V102 V103 V104 V105 V106 V107 V108
                                       V109 V110 V111 V112 V113 V114 V115 V116 V117 V118 V119 V120 V121 V122 V1
23 V124 V125 V126 V127 V128 V129 V130 V131 V132 V133 V134 V135 V136 V137 V138 V139 V140 V141 V142 V143 V144 V14
5 V146 V147 V148 V149 V150 V151
                                       V152 V153 V154 V155 V156 V157 V158 V159 V160 V161 V162 V163 V164 V165 V1
66 V167 V168 V169 V170 V171 V172 V173 V174 V175 V176 V177 V178 V179 V180 V181 V182 V183 V184 V185 V186 V187 V18
8 V189 V190 V191 V192 V193 V194
                                       V195 V196 V197 V198 V199 V200 V201 V202 V203 V204 V205 V206 V207 V208 V2
09 V210 V211 V212 V213 V214 V215 V216 V217 V218 V219 V220 V221 V222 V223 V224 V225 V226 V227 V228 V229 V230 V23
1 V232
Selected controls (12):                V2 V3 V6 V8 V9 V10 V11 V12 V13 V14 V15 V25
Unpenalized controls (1):              _cons

Structural equation:

OLS using CHS lasso-orthogonalized vars
------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           d |   1.841995   .0949758    19.39   0.000     1.655846    2.028145
------------------------------------------------------------------------------

OLS using CHS post-lasso-orthogonalized vars
------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           d |   1.443783    .151316     9.54   0.000     1.147209    1.740357
------------------------------------------------------------------------------

OLS with PDS-selected variables and full regressor set
------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           d |   1.443783   .1244193    11.60   0.000     1.199926    1.687641
          V2 |   .0304683   .3224507     0.09   0.925    -.6015234      .66246
          V3 |   .0200068   .3141448     0.06   0.949    -.5957056    .6357192
          V6 |   .1951426   .3118275     0.63   0.531    -.4160281    .8063133
          V8 |   .6232437   .2980334     2.09   0.037     .0391091    1.207378
          V9 |   .8211188   .3155202     2.60   0.009     .2027105    1.439527
         V10 |   .5466191   .3027309     1.81   0.071    -.0467227    1.139961
         V11 |   .6675631   .2943451     2.27   0.023     .0906573    1.244469
         V12 |   .7539879   .3097648     2.43   0.015     .1468599    1.361116
         V13 |   .3324538   .3340122     1.00   0.320    -.3221982    .9871058
         V14 |   .4096044    .297479     1.38   0.169    -.1734437    .9926524
         V15 |      .3454   .3314432     1.04   0.297    -.3042167    .9950166
         V25 |   1.149046   .2918479     3.94   0.000     .5770349    1.721057
       _cons |   .5076797   .3145461     1.61   0.107    -.1088195    1.124179
------------------------------------------------------------------------------
Standard errors and test statistics valid for the following variables only:
    d
------------------------------------------------------------------------------

.
```

Suppose that we don't have information on potential confounders(V1-V15), only the potential IVs (V16-V232). Using Lasso to do the first step.

In [11]:
```
%%stata
capture drop d_hat

quietly lasso2 d V31-V232,  lic(aicc) postres
predict d_hat, xb

reg y V16-V30 d_hat

ivregress 2sls y V16-V30 (d=V31-V32)

drop d_hat

qui rlasso d V31-V232
predict d_hat, xb
reg y V16-V30 d_hat

drop d_hat

qui cvlasso d V31-V232, lopt  postres
predict d_hat, xb
reg y V16-V30 d_hat
```

```
. capture drop d_hat

.
. quietly lasso2 d V31-V232,  lic(aicc) postres

. predict d_hat, xb
Use e(b) from previous lasso2 estimation (lambda=152.4505926358).

.
. reg y V16-V30 d_hat

      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(16, 183)      =     10.93
       Model |  6921.19811        16  432.574882   Prob > F        =    0.0000
    Residual |  7242.21108       183   39.574924   R-squared       =    0.4887
-------------+----------------------------------   Adj R-squared   =    0.4440
       Total |  14163.4092       199  71.1729105   Root MSE        =    6.2909

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         V16 |    1.68601    .460915     3.66   0.000     .7766188    2.595401
         V17 |   .8359224   .4570133     1.83   0.069    -.0657702    1.737615
         V18 |   .9915455    .484952     2.04   0.042     .0347295    1.948362
         V19 |   .2383975   .4534881     0.53   0.600    -.6563399    1.133135
         V20 |   1.025299   .4751977     2.16   0.032     .0877287     1.96287
         V21 |    1.14743   .4609464     2.49   0.014     .2379775    2.056883
         V22 |   .4570792    .460582     0.99   0.322    -.4516545    1.365813
         V23 |   .4667879   .4448285     1.05   0.295    -.4108641     1.34444
         V24 |    .878771   .4518032     1.95   0.053     -.012642    1.770184
         V25 |   .4082248   .4571781     0.89   0.373    -.4937931    1.310243
         V26 |   .9238036    .466177     1.98   0.049     .0040309    1.843576
         V27 |   1.066985   .4574228     2.33   0.021     .1644846    1.969486
         V28 |   .7401802    .441445     1.68   0.095     -.130796    1.611156
         V29 |   .9312996   .4436147     2.10   0.037     .0560424    1.806557
         V30 |   .1744806   .4518461     0.39   0.700    -.7170172    1.065978
       d_hat |   4.081889   .3804974    10.73   0.000     3.331163    4.832614
       _cons |  -1.799789   .5872856    -3.06   0.003    -2.958511   -.6410675
------------------------------------------------------------------------------

.
. ivregress 2sls y V16-V30 (d=V31-V32)

Instrumental variables 2SLS regression            Number of obs   =        200
                                                  Wald chi2(16)   =      82.00
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.5833
                                                  Root MSE        =     5.4322

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           d |   .5892811   .4398329     1.34   0.180    -.2727756    1.451338
         V16 |   .9644799   .3966995     2.43   0.015     .1869632    1.741997
         V17 |   .6063748   .4097583     1.48   0.139    -.1967366    1.409486
         V18 |   .7713051   .4352889     1.77   0.076    -.0818454    1.624456
         V19 |   .5386584   .4201282     1.28   0.200    -.2847777    1.362095
         V20 |   1.133044   .4201642     2.70   0.007     .3095368     1.95655
         V21 |   1.527868    .434471     3.52   0.000     .6763204    2.379416
         V22 |   .8292419   .3993792     2.08   0.038      .046473    1.612011
         V23 |   .4808632   .3897115     1.23   0.217    -.2829573    1.244684
         V24 |   .9584635   .3940326     2.43   0.015     .1861738    1.730753
         V25 |  -.2622715   .5707806    -0.46   0.646    -1.380981     .856438
         V26 |   .6087604   .4422242     1.38   0.169    -.2579831    1.475504
         V27 |   1.426428   .4099524     3.48   0.001      .622936     2.22992
         V28 |   .6518562   .4075931     1.60   0.110    -.1470117    1.450724
         V29 |    .549518   .4058813     1.35   0.176    -.2459946    1.345031
         V30 |   .4970616    .439328     1.13   0.258    -.3640055    1.358129
       _cons |   1.612411   .5864586     2.75   0.006     .4629731    2.761849
------------------------------------------------------------------------------
Instrumented: d
 Instruments: V16 V17 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30
              V31 V32

.
. drop d_hat

.
. qui rlasso d V31-V232

. predict d_hat, xb

. reg y V16-V30 d_hat
note: d_hat omitted because of collinearity.

      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(15, 184)      =      2.46
       Model |  2366.71528        15  157.781019   Prob > F        =    0.0026
    Residual |  11796.6939       184  64.1124669   R-squared       =    0.1671
-------------+----------------------------------   Adj R-squared   =    0.0992
       Total |  14163.4092       199  71.1729105   Root MSE        =     8.007

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         V16 |   .8903259   .5790085     1.54   0.126    -.2520234    2.032675
         V17 |   .4529499   .5799105     0.78   0.436     -.691179    1.597079
         V18 |   .6068155   .6155582     0.99   0.326    -.6076443    1.821275
         V19 |   .3344905   .5770884     0.58   0.563    -.8040706    1.473052
         V20 |   1.256609   .6042101     2.08   0.039     .0645381    2.448679
         V21 |   1.770703    .582015     3.04   0.003      .622422    2.918984
         V22 |   .7699611   .5850536     1.32   0.190    -.3843149    1.924237
         V23 |    .392334   .5661102     0.69   0.489    -.7245678    1.509236
         V24 |   .8844222    .575056     1.54   0.126    -.2501292    2.018974
         V25 |  -.8305067   .5630318    -1.48   0.142    -1.941335    .2803214
         V26 |   .3559534   .5895141     0.60   0.547    -.8071229     1.51903
         V27 |   1.583662   .5789728     2.74   0.007     .4413829    2.725941
         V28 |   .4561087   .5608607     0.81   0.417    -.6504362    1.562654
         V29 |   .3594885   .5605436     0.64   0.522    -.7464307    1.465408
         V30 |   .2264722    .575078     0.39   0.694    -.9081226    1.361067
       d_hat |          0  (omitted)
       _cons |   2.196782   .5778513     3.80   0.000     1.056716    3.336848
------------------------------------------------------------------------------

.
. drop d_hat

.
. qui cvlasso d V31-V232, lopt  postres
Warning: lse is at the limit of the lambda range.

. predict d_hat, xb
Use e(b) from previous lasso2 estimation (lambda=152.4505926358).

. reg y V16-V30 d_hat

      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(16, 183)      =     10.93
       Model |  6921.19811        16  432.574882   Prob > F        =    0.0000
    Residual |  7242.21108       183   39.574924   R-squared       =    0.4887
-------------+----------------------------------   Adj R-squared   =    0.4440
       Total |  14163.4092       199  71.1729105   Root MSE        =    6.2909

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         V16 |    1.68601    .460915     3.66   0.000     .7766188    2.595401
         V17 |   .8359224   .4570133     1.83   0.069    -.0657702    1.737615
         V18 |   .9915455    .484952     2.04   0.042     .0347295    1.948362
         V19 |   .2383975   .4534881     0.53   0.600    -.6563399    1.133135
         V20 |   1.025299   .4751977     2.16   0.032     .0877287     1.96287
         V21 |    1.14743   .4609464     2.49   0.014     .2379775    2.056883
         V22 |   .4570792    .460582     0.99   0.322    -.4516545    1.365813
         V23 |   .4667879   .4448285     1.05   0.295    -.4108641     1.34444
         V24 |    .878771   .4518032     1.95   0.053     -.012642    1.770184
         V25 |   .4082248   .4571781     0.89   0.373    -.4937931    1.310243
         V26 |   .9238036    .466177     1.98   0.049     .0040309    1.843576
         V27 |   1.066985   .4574228     2.33   0.021     .1644846    1.969486
         V28 |   .7401802    .441445     1.68   0.095     -.130796    1.611156
         V29 |   .9312996   .4436147     2.10   0.037     .0560424    1.806557
         V30 |   .1744806   .4518461     0.39   0.700    -.7170172    1.065978
       d_hat |   4.081889   .3804974    10.73   0.000     3.331163    4.832614
       _cons |  -1.799789   .5872856    -3.06   0.003    -2.958511   -.6410675
------------------------------------------------------------------------------

.
```

The IVlasso and the kitchen sink version of IV

In [12]:

```
%%stata
ivregress 2sls y V16-V30 (d=V31-V232)
# matrix list e(b)

ivlasso y V16-V30 (d=V31-V232)
# matrix list e(b)
```

```
. ivregress 2sls y V16-V30 (d=V31-V232)
note: V230 omitted because of collinearity.
note: V231 omitted because of collinearity.
note: V232 omitted because of collinearity.
note: V214 omitted because of collinearity.
note: V215 omitted because of collinearity.
note: V216 omitted because of collinearity.
note: V217 omitted because of collinearity.
note: V218 omitted because of collinearity.
note: V219 omitted because of collinearity.
note: V220 omitted because of collinearity.
note: V221 omitted because of collinearity.
note: V222 omitted because of collinearity.
note: V223 omitted because of collinearity.
note: V224 omitted because of collinearity.
note: V225 omitted because of collinearity.
note: V226 omitted because of collinearity.
note: V227 omitted because of collinearity.
note: V228 omitted because of collinearity.
note: V229 omitted because of collinearity.

Instrumental variables 2SLS regression            Number of obs   =        200
                                                  Wald chi2(16)   =    4275.32
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.9553
                                                  Root MSE        =      1.779

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           d |   1.883095   .0317059    59.39   0.000     1.820953    1.945238
         V16 |   1.127291   .1287033     8.76   0.000      .875037    1.379545
         V17 |   .9432317   .1291061     7.31   0.000     .6901885    1.196275
         V18 |   1.132455    .137048     8.26   0.000     .8638461    1.401064
         V19 |   .9869256   .1286846     7.67   0.000     .7347084    1.239143
         V20 |   .8617462   .1344052     6.41   0.000     .5983169    1.125176
         V21 |   .9947043   .1299679     7.65   0.000     .7399719    1.249437
         V22 |   .9593975   .1300237     7.38   0.000     .7045558    1.214239
         V23 |   .6752362    .125866     5.36   0.000     .4285434     .921929
         V24 |   1.121027   .1278254     8.77   0.000     .8704939     1.37156
         V25 |   .9853348   .1287739     7.65   0.000     .7329426    1.237727
         V26 |   1.163819     .13168     8.84   0.000     .9057306    1.421907
         V27 |   1.081208   .1289115     8.39   0.000      .828546     1.33387
         V28 |   1.081636   .1250538     8.65   0.000     .8365347    1.326737
         V29 |   .9667432    .124958     7.74   0.000       .72183    1.211656
         V30 |   1.091162    .128595     8.49   0.000     .8391208    1.343204
       _cons |   .3293765   .1321784     2.49   0.013     .0703115    .5884414
------------------------------------------------------------------------------
Instrumented: d
 Instruments: V16 V17 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30
              V31 V32 V33 V34 V35 V36 V37 V38 V39 V40 V41 V42 V43 V44 V45
              V46 V47 V48 V49 V50 V51 V52 V53 V54 V55 V56 V57 V58 V59 V60
              V61 V62 V63 V64 V65 V66 V67 V68 V69 V70 V71 V72 V73 V74 V75
              V76 V77 V78 V79 V80 V81 V82 V83 V84 V85 V86 V87 V88 V89 V90
              V91 V92 V93 V94 V95 V96 V97 V98 V99 V100 V101 V102 V103 V104
              V105 V106 V107 V108 V109 V110 V111 V112 V113 V114 V115 V116
              V117 V118 V119 V120 V121 V122 V123 V124 V125 V126 V127 V128
              V129 V130 V131 V132 V133 V134 V135 V136 V137 V138 V139 V140
              V141 V142 V143 V144 V145 V146 V147 V148 V149 V150 V151 V152
              V153 V154 V155 V156 V157 V158 V159 V160 V161 V162 V163 V164
              V165 V166 V167 V168 V169 V170 V171 V172 V173 V174 V175 V176
              V177 V178 V179 V180 V181 V182 V183 V184 V185 V186 V187 V188
              V189 V190 V191 V192 V193 V194 V195 V196 V197 V198 V199 V200
              V201 V202 V203 V204 V205 V206 V207 V208 V209 V210 V211 V212
              V213

. # matrix list e(b)
Unknown #command
.
. ivlasso y V16-V30 (d=V31-V232)
5.  (PDS/CHS) Selecting HD controls/IVs for endog regressor d...
Selected:
7.  (CHS) Creating orthogonalized endogenous regressor d...


Estimation results:

Specification:
Regularization method:                   lasso
Penalty loadings:                        homoskedastic
Number of observations:                  200
Exogenous (15):                          V16 V17 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30
Endogenous (1):                          d
Unpenalized controls (1):                _cons
High-dim instruments (202):              V31 V32 V33 V34 V35 V36 V37 V38 V39 V40 V41 V42 V43 V44 V45 V46 V47 V48
V49 V50 V51 V52 V53 V54 V55 V56 V57 V58 V59 V60 V61 V62 V63 V64 V65 V66 V67 V68 V69 V70 V71 V72 V73 V74 V75 V76
V77 V78 V79 V80 V81 V82 V83 V84
                                         V85 V86 V87 V88 V89 V90 V91 V92 V93 V94 V95 V96 V97 V98 V99 V100 V101 V1
02 V103 V104 V105 V106 V107 V108 V109 V110 V111 V112 V113 V114 V115 V116 V117 V118 V119 V120 V121 V122 V123 V12
4 V125 V126 V127 V128 V129 V130
                                         V131 V132 V133 V134 V135 V136 V137 V138 V139 V140 V141 V142 V143 V144 V1
45 V146 V147 V148 V149 V150 V151 V152 V153 V154 V155 V156 V157 V158 V159 V160 V161 V162 V163 V164 V165 V166 V16
7 V168 V169 V170 V171 V172 V173
                                         V174 V175 V176 V177 V178 V179 V180 V181 V182 V183 V184 V185 V186 V187 V1
88 V189 V190 V191 V192 V193 V194 V195 V196 V197 V198 V199 V200 V201 V202 V203 V204 V205 V206 V207 V208 V209 V21
0 V211 V212 V213 V214 V215 V216
                                         V217 V218 V219 V220 V221 V222 V223 V224 V225 V226 V227 V228 V229 V230 V2
31 V232
Selected instruments (0):

Structural equation:

IV using CHS lasso-orthogonalized vars
------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           d |   2.215239   .1611988    13.74   0.000     1.899295    2.531183
         V16 |   1.169087   .1635883     7.15   0.000     .8484601    1.489714
         V17 |   1.029709   .1663215     6.19   0.000     .7037243    1.355693
         V18 |   1.225169   .1779893     6.88   0.000     .8763161    1.574021
         V19 |   1.102003   .1688166     6.53   0.000     .7711289    1.432878
         V20 |   .7920996    .171461     4.62   0.000     .4560422    1.128157
         V21 |   .8578322   .1714585     5.00   0.000     .5217798    1.193885
         V22 |   .9928107   .1629563     6.09   0.000     .6734222    1.312199
         V23 |    .725135   .1608674     4.51   0.000     .4098408    1.040429
         V24 |    1.16276   .1587704     7.32   0.000     .8515758    1.473944
         V25 |   1.305616   .2173506     6.01   0.000      .879617    1.731616
         V26 |   1.306311   .1784873     7.32   0.000     .9564827     1.65614
         V27 |    .992584   .1678479     5.91   0.000     .6636081     1.32156
         V28 |   1.191967   .1629848     7.31   0.000      .872523    1.511412
         V29 |   1.073852    .166811     6.44   0.000     .7469084    1.400796
         V30 |   1.243678   .1738472     7.15   0.000     .9029438    1.584412
------------------------------------------------------------------------------

IV using CHS post-lasso-orthogonalized vars
------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           d |   2.215239   .1611988    13.74   0.000     1.899295    2.531183
         V16 |   1.169087   .1635883     7.15   0.000     .8484601    1.489714
         V17 |   1.029709   .1663215     6.19   0.000     .7037243    1.355693
         V18 |   1.225169   .1779893     6.88   0.000     .8763161    1.574021
         V19 |   1.102003   .1688166     6.53   0.000     .7711289    1.432878
         V20 |   .7920996    .171461     4.62   0.000     .4560422    1.128157
         V21 |   .8578322   .1714585     5.00   0.000     .5217798    1.193885
         V22 |   .9928107   .1629563     6.09   0.000     .6734222    1.312199
         V23 |    .725135   .1608674     4.51   0.000     .4098408    1.040429
         V24 |    1.16276   .1587704     7.32   0.000     .8515758    1.473944
         V25 |   1.305616   .2173506     6.01   0.000      .879617    1.731616
         V26 |   1.306311   .1784873     7.32   0.000     .9564827     1.65614
         V27 |    .992584   .1678479     5.91   0.000     .6636081     1.32156
         V28 |   1.191967   .1629848     7.31   0.000      .872523    1.511412
         V29 |   1.073852    .166811     6.44   0.000     .7469084    1.400796
         V30 |   1.243678   .1738472     7.15   0.000     .9029438    1.584412
------------------------------------------------------------------------------

IV with PDS-selected variables and full regressor set
------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           d |          0  (omitted)
         V16 |          0  (omitted)
         V17 |          0  (omitted)
         V18 |          0  (omitted)
         V19 |          0  (omitted)
         V20 |          0  (omitted)
         V21 |          0  (omitted)
         V22 |          0  (omitted)
         V23 |          0  (omitted)
         V24 |          0  (omitted)
         V25 |          0  (omitted)
         V26 |          0  (omitted)
         V27 |          0  (omitted)
         V28 |          0  (omitted)
         V29 |          0  (omitted)
         V30 |          0  (omitted)
         V30 |          0  (omitted)
------------------------------------------------------------------------------
Standard errors and test statistics valid for the following variables only:
    d V16 V17 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30
------------------------------------------------------------------------------
Warning: not enough instruments selected; model unidentified

. # matrix list e(b)
Unknown #command
.
```

Distribution of estimators

Run the simulation for 100 times and look at the distribution of the different estimators.

In [ ]:

```
%%stata

clear all

set seed 20211

capture postutil clear
tempfile simresults
postfile sim_mem  OVB OLS LASSO2 PDS DS P_out IV IVlasso using "`simresults'", replace

forvalues i=1/100 {

        quietly set obs 5000

        di `i'


    forvalues k=1/232{
            g V`k'=rnormal(0,1)
        }

    egen sum_c=rowtotal(V1-V15)
    egen sum_y=rowtotal(V16-V30)
    egen sum_e=rowtotal(V31-V32)

    g d=1+1*sum_c+sum_e+rnormal()
    g y=1+d+sum_c+sum_y+rnormal()

        * OVB
        quietly reg y d
        mat R=e(b)

        * DGR OLS
        quietly reg y d V1-V30
        mat A=e(b)

    *lasso2
        quietly lasso2 y d V*, lic(aicc) postres
        mat B=e(beta)


        ** own post double selection using lasso2
        quietly lasso2 y V*, lic(aicc) postres
        local string1 = e(selected)

        quietly lasso2 d V*, lic(aicc) postres
        local string2 = e(selected)

        local newlist: list string1 | string2
        * di "`newlist'"

        quietly reg y d `newlist'
        mat F=e(b)

        ** packaged
        quietly pdslasso y d (V*)
        mat D=e(beta_pds)

    ** Partialling out
    quietly reg y `string1'
    predict y_dot, res

    quietly reg d `string2'
    predict d_dot, res

    quietly reg y_dot d_dot

    mat G=e(b)

    ** IV
    quietly ivregress 2sls y V16-V30 (d=V31-V232)
    mat IV1=e(b)

    ** IVlasso
    quietly ivlasso y V16-V30 (d=V31-V232)
    mat IV2=e(b)

        post sim_mem (R[1,1]) (A[1,1]) (B[1,1]) (D[1,1]) (F[1,1]) (G[1,1]) (IV1[1,1]) (IV2[1,1])

        capture drop y d V* sum* *_dot
}
postclose sim_mem

use "`simresults'", clear

save sim1_1_5000.dta, replace
```

case I: 200 obs vs 5000 obs, all parameters equal to 1.

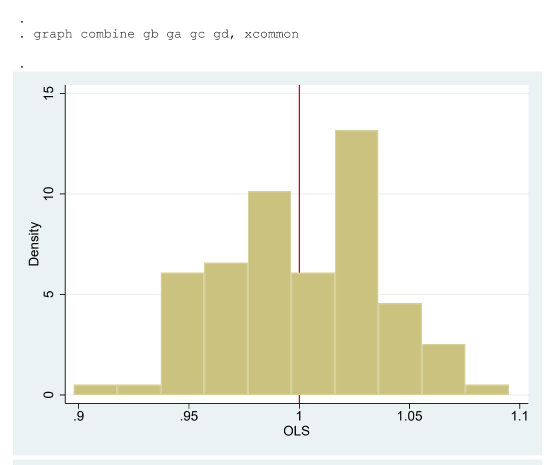
$$y = \alpha d + \sum_{k=1}^{K_c} \beta_k^{cy} x_k^c + \sum_{k=1}^{K_y} \beta_k^y x_k^y + \varepsilon^y$$

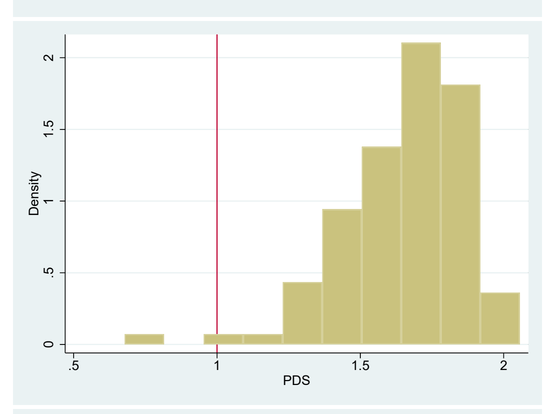
$$d = \sum_{k=1}^{K_c} \beta_k^{cd} x_k^c + \sum_{k=1}^{K_e} \beta_k^e x_k^e + \varepsilon^d$$

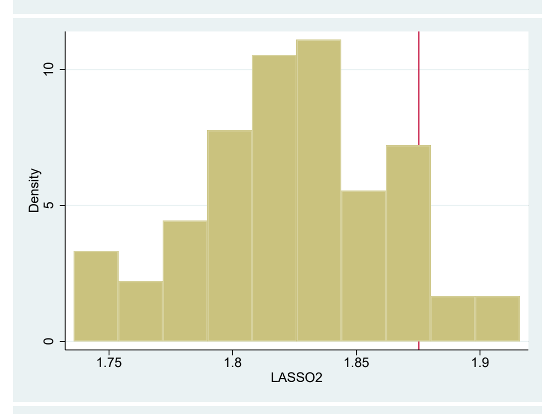
In [13]:

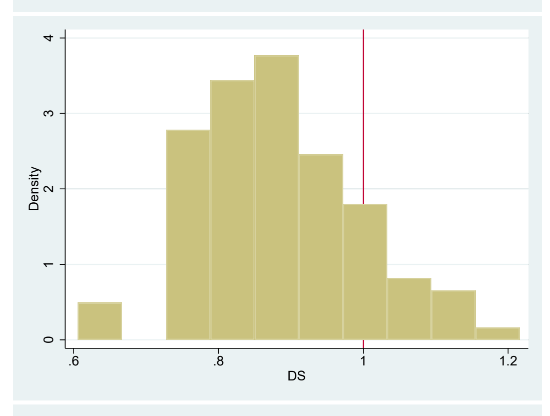
```
%%stata
clear all

use sim1_1_200.dta, clear

su

qui {
    twoway hist OLS , xline(1.0) name(ga)
    twoway hist PDS, xline(1.0) name(gb)
    twoway hist LASSO2, xline(1.0) name(gc)
    twoway hist DS, xline(1.0) name(gd)
}

graph combine gb ga gc gd, xcommon
```

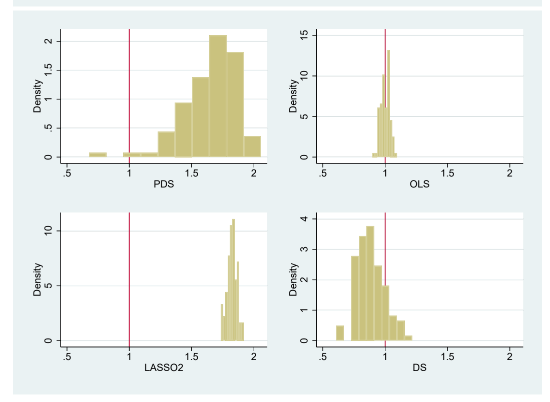
```
. clear all

.
. use sim1_1_200.dta, clear

.
. su

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         OVB |        100    1.850334    .0715134   1.689236   2.002945
         OLS |        100    1.001209    .0374376   .8977358   1.095102
      LASSO2 |        100    1.824524    .0382321   1.735735   1.916057
         PDS |        100    1.655872     .218394   .6774625   2.056151
          DS |        100    .8869708    .1114914   .6057009   1.216065
-------------+---------------------------------------------------------
       P_out |        100    .7294227    .1274495   .4364092   1.146187
          IV |        100    1.834283    .0352733   1.754402   1.906261
     IVlasso |        100    .5022424    .6092893          0   1.515303

.
. qui {

.
. graph combine gb ga gc gd, xcommon

.
```







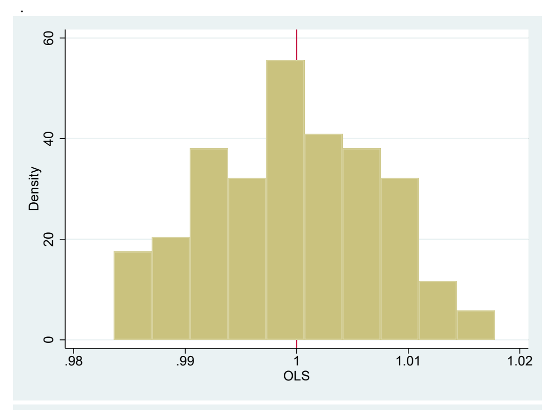


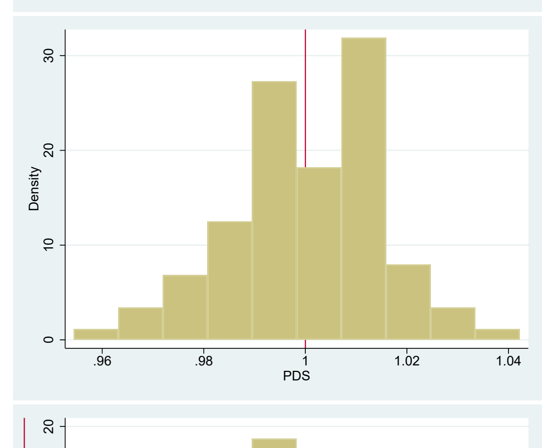


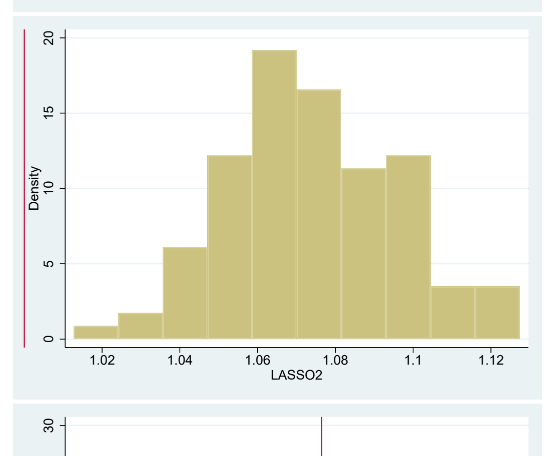
In [14]:

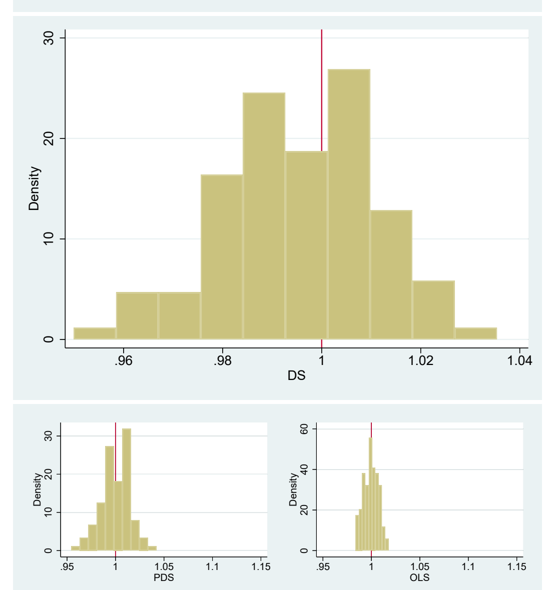
```
%%stata
clear all

use sim1_1_5000.dta, clear

su

qui {
    twoway hist OLS , xline(1.0) name(ga)
    twoway hist PDS, xline(1.0) name(gb)
    twoway hist LASSO2, xline(1.0) name(gc)
    twoway hist DS, xline(1.0) name(gd)
}

graph combine gb ga gc gd, xcommon
```

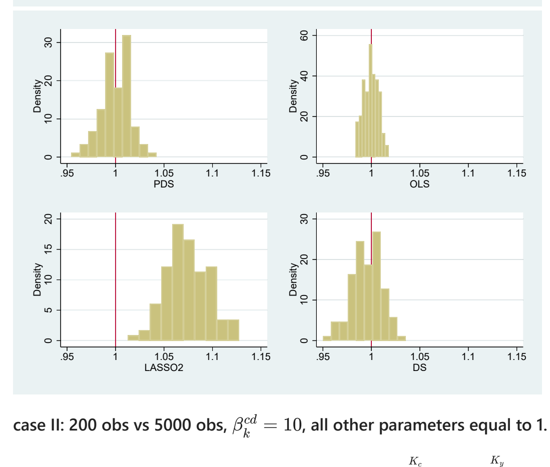
```
. clear all

.
. use sim1_1_5000.dta, clear

.
. su

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         OVB |        100    1.834414    .0156417   1.785204   1.867691
         OLS |        100    .9995977    .0076402   .9836014   1.017771
      LASSO2 |        100    1.074105    .0222739   1.012695   1.127418
         PDS |        100     1.00037    .0148173   .9543703   1.042218
          DS |        100    .9952539    .0152882   .9499198   1.035413
-------------+---------------------------------------------------------
       P_out |        100    .9895131    .0160422   .9427422   1.027128
          IV |        100    1.228091    .0270834     1.1654   1.291337
     IVlasso |        100    .9994752    .0387969   .9173788   1.081761

.
. qui {

.
. graph combine gb ga gc gd, xcommon

.
```











case II: 200 obs vs 5000 obs, $\beta_k^{cd} = 10$, all other parameters equal to 1.

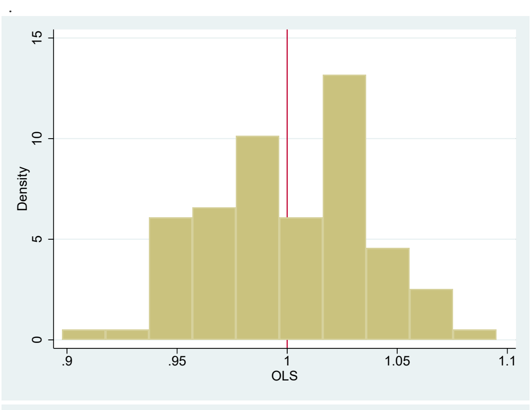
$$y = 1 + \alpha d + \sum_{k=1}^{K_c} \beta_k^{cy} x_k^c + \sum_{k=1}^{K_y} \beta_k^y x_k^y + \varepsilon^y$$

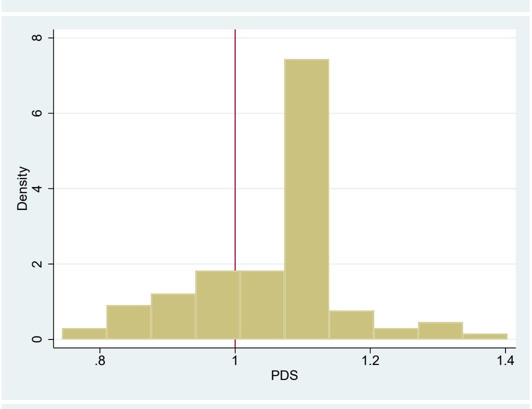
$$d = 1 + \sum_{k=1}^{K_c} 10 x_k^c + \sum_{k=1}^{K_e} \beta_k^e x_k^e + \varepsilon^d$$

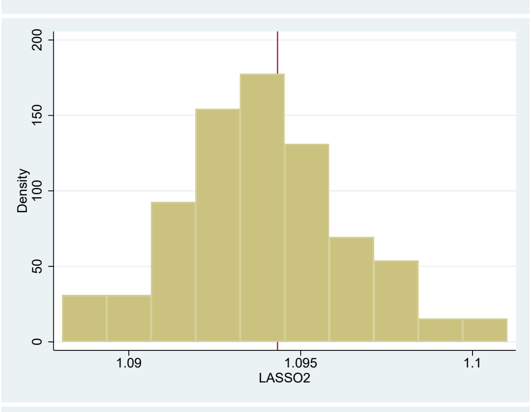
In [15]:

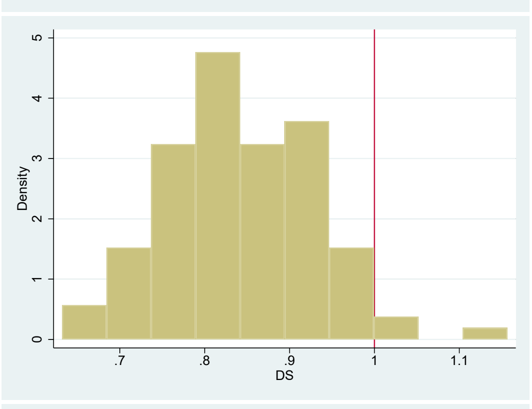
```
%%stata
clear all

use sim1_10_200.dta, clear

su

twoway hist OLS , xline(1.0) name(ga)

twoway hist PDS, xline(1.0) name(gb)

twoway hist LASSO2, xline(1.0) name(gc)

twoway hist DS, xline(1.0) name(gd)

graph combine gb ga gc gd, xcommon
```

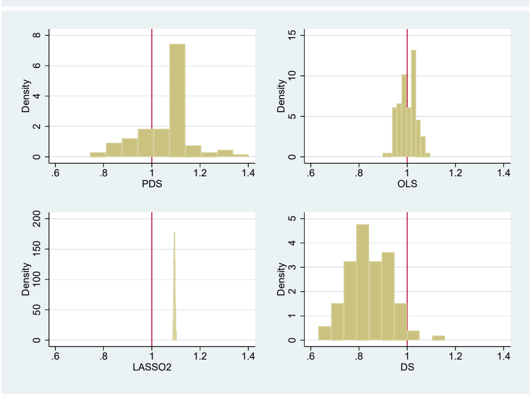
```
. clear all

.
. use sim1_10_200.dta, clear

.
. su

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         OVB |        100    1.101343    .0070753   1.080156   1.119817
         OLS |        100    1.001209    .0374376   .8977359   1.095103
      LASSO2 |        100     1.09386    .0024848   1.088058   1.101014
         PDS |        100    1.059699    .1101767   .7438613   1.402718
          DS |        100    .8432862    .0895669   .6320081   1.156564
-------------+---------------------------------------------------------
       P_out |        100    .6107089    .0921731   .4088039    .978497
          IV |        100    1.100013     .001968   1.093904   1.103477
     IVlasso |        100           0           0          0          0

.
. twoway hist OLS , xline(1.0) name(ga)

.
. twoway hist PDS, xline(1.0) name(gb)

.
. twoway hist LASSO2, xline(1.0) name(gc)

.
. twoway hist DS, xline(1.0) name(gd)

.
. graph combine gb ga gc gd, xcommon

.
```







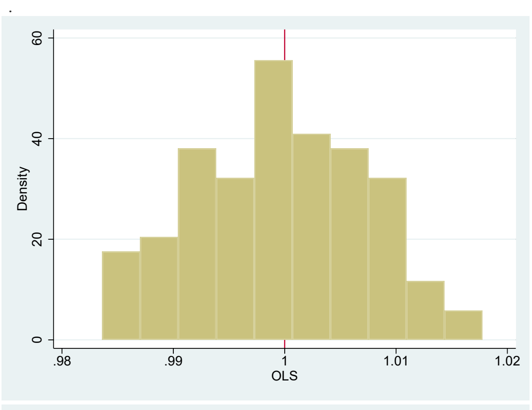


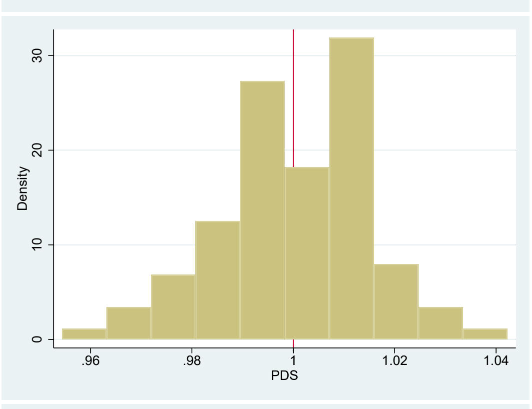


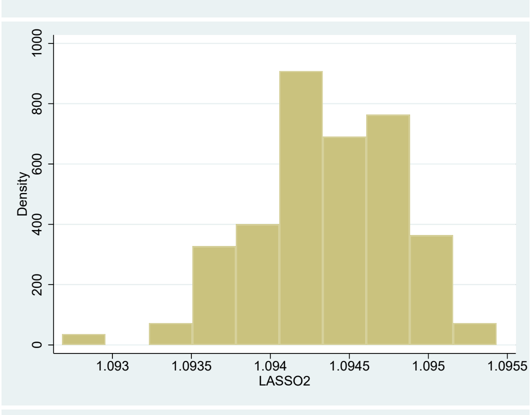
In [16]:

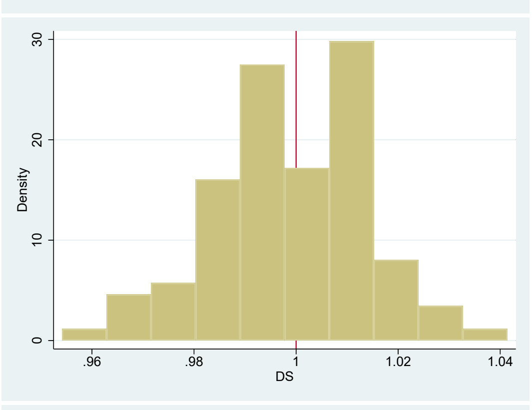
```
%%stata
clear all

use sim1_10_5000.dta, clear

su

twoway hist OLS , xline(1.0) name(ga)

twoway hist PDS, xline(1.0) name(gb)

twoway hist LASSO2, xline(1.0) name(gc)

twoway hist DS, xline(1.0) name(gd)

graph combine gb ga gc gd, xcommon
```

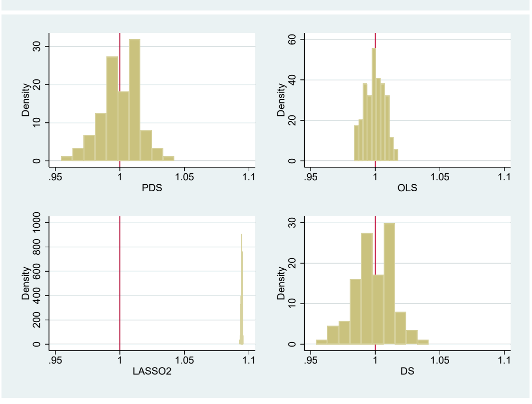
```
. clear all

.
. use sim1_10_5000.dta, clear

.
. su

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         OVB |        100    1.099956    .0015691   1.095716   1.104368
         OLS |        100    .9995977    .0076402   .9836013   1.017771
      LASSO2 |        100    1.094373    .0004657   1.092681   1.095432
         PDS |        100    1.000398    .0148331   .9543703   1.042218
          DS |        100    .9990506    .0149462   .9541339   1.041394
-------------+---------------------------------------------------------
       P_out |        100    .9941824    .0154224    .952083   1.036947
          IV |        100    1.096747    .0020795    1.09095   1.102292
     IVlasso |        100     .010633    .1063295          0   1.063295

.
. twoway hist OLS , xline(1.0) name(ga)

.
. twoway hist PDS, xline(1.0) name(gb)

.
. twoway hist LASSO2, xline(1.0) name(gc)

.
. twoway hist DS, xline(1.0) name(gd)

.
. graph combine gb ga gc gd, xcommon

.
```











Sometimes Double Selection might not be better than Naive Lasso

- more likely to mistakenly include valid IV variables into the model when using DS
- less likely to drop counfounders in Naive Lasso
- sparsity condition! Now we have 70 valid IV ($V31 - V100$) for D.

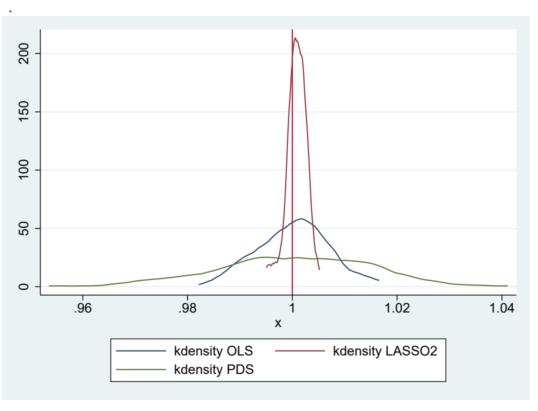
In [ ]:

```
%%stata

clear all

set seed 20211

capture postutil clear
tempfile simresults
postfile sim_mem  OVB OLS LASSO2 PDS using "`simresults'", replace

forvalues i=1/100 {

        quietly set obs 200

        * di `i'

    su

    forvalues k=1/232{
            g V`k'=rnormal(0,1)
        }

    egen sum_c=rowtotal(V1-V15)
    egen sum_y=rowtotal(V16-V30)
    egen sum_e=rowtotal(V31-V100)

    g d=1+sum_c+sum_e+rnormal()
    g y=1+d+sum_c+sum_y+rnormal()

        * OVB
        quietly reg y d
        mat R=e(b)

        * OLS
        quietly reg y d V1-V15
        mat A=e(b)

        quietly lasso2 y d V*
        quietly lasso2, lic(aicc) postres
        mat B=e(betaOLS)


        ** packaged
        quietly pdslasso y d (V*)
        mat D=e(beta_pds)


        post sim_mem (R[1,1]) (A[1,1]) (B[1,1]) (D[1,1])

    capture drop y d V* sum*
}
postclose sim_mem

use "`simresults'", clear
save sim2_200.dta, replace
```

In [17]:

```
%%stata

# 5000 observations

use sim2, clear

twoway kdensity OLS ||kdensity LASSO2 || kdensity PDS, xline(1.0)

su OVB OLS LASSO2 PDS
```

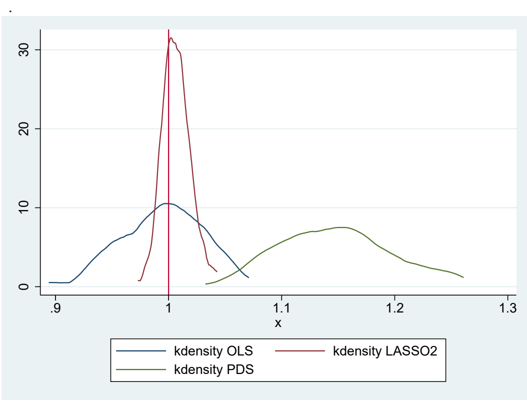
```
.
. use sim2, clear

.
. twoway kdensity OLS ||kdensity LASSO2 || kdensity PDS, xline(1.0)

.
. su OVB OLS LASSO2 PDS

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         OVB |        100    1.173633    .0089582   1.153944    1.19557
         OLS |        100    1.000203    .0068805   .9821414   1.016598
      LASSO2 |        100    1.000835    .0019208    .995064   1.005203
         PDS |        100    1.000131    .0148838   .9535071   1.041129

.
```



In [18]:

```
%%stata

use sim2_200, clear

twoway kdensity OLS ||kdensity LASSO2 || kdensity PDS, xline(1.0)

su OVB OLS LASSO2 PDS
```

```
.
. use sim2_200, clear

.
. twoway kdensity OLS ||kdensity LASSO2 || kdensity PDS, xline(1.0)

.
. su OVB OLS LASSO2 PDS

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         OVB |        100    1.169192    .0440991   1.081015   1.260978
         OLS |        100    .9947198    .0360946    .894028   1.071052
      LASSO2 |        100    1.006816    .0124615   .9728466   1.042922
         PDS |        100    1.147914    .0488963   1.032479   1.260978

.
```



In [ ]: