hdm_intro_r

October 5, 2022

1 Hight-Dimmensional Metrics in Julia

2 Introduction

3 How to Get Started

[57]:
```
install.packages("hdm")
```

```
Warning message:
"package 'hdm' is in use and will not be installed"
```

4 Prediction Using Approximate Sparsity

[58]:
```
library(hdm)
# 3.2 A Joint Significance test for Lasso Regression
set.seed(12345)
n = 100
#sample size
p = 50
# number of variables
s = 3
# nubmer of variables with non-zero coefficients
X = matrix(rnorm(n * p), ncol = p)
beta = c(rep(5, s), rep(0, p - s))
Y = X %*% beta + rnorm(n)

# cbind(Y, X)

r_data = function (name = "nn", ...){
    nn = paste0("r_", name, ".csv")
    # cbind(...)
    dta = cbind(...)
    write.csv(dta, nn, row.names = F)
    print(head(dta))
}

# r_data()
```

```r
r_data(name = "01", Y, X)
```

```
           [,1]       [,2]       [,3]       [,4]        [,5]         [,6]
[1,]  -3.825930  0.5855288  0.2239254 -1.4361457  0.52228217  0.627965113
[2,]  -4.255389  0.7094660 -1.1562233 -0.6292596  0.00979376  0.002143951
[3,]   2.957720 -0.1093033  0.4224185  0.2435218 -0.44052620  0.284377723
[4,]  -5.567258 -0.4534972 -1.3247553  1.0583622  1.19948953 -1.001779086
[5,]   6.247331  0.6058875  0.1410843  0.8313488 -0.11746849 -0.617221929
[6,] -11.353745 -1.8179560 -0.5360480  0.1052118  0.03820979  0.828194239
           [,7]       [,8]        [,9]      [,10]       [,11]       [,12]
[1,] -1.4203239 -1.6366291  2.30362701 -0.8174921 -0.78486098  1.67751179
[2,] -2.4669386  0.2115626  2.02089590 -0.2492659 -2.56005244  0.07947405
[3,]  0.4847158 -0.4648317 -0.05787852  0.4629986  0.07280078 -0.85642750
[4,] -0.9379723 -0.6623572  0.44209338  0.6673264  0.75024358 -0.77877729
[5,]  3.3307333 -0.1329536 -1.30333114  0.4881699 -0.12824888 -0.38093608
[6,] -0.1629455 -1.3217017 -0.03522043  1.0764874 -0.48786673 -1.89735834
          [,13]       [,14]      [,15]       [,16]      [,17]      [,18]
[1,]  1.1480914 -0.26788171 -1.6598937 -0.76455739 -0.1746226  0.7479930
[2,]  0.4550137  0.93728801 -0.2763602  0.97919595 -0.6706167  1.3071286
[3,]  2.0219464  0.03402358 -0.8284337  1.17383262  0.5074258 -0.5878296
[4,] -0.6952704  0.24315886 -0.6860709  1.37466642  1.2474343 -1.1335791
[5,]  1.6039653  1.14368852 -0.2716078  1.20273835 -1.2482755 -0.7816046
[6,]  1.3517073  0.82740593  1.2438106  0.02694937 -1.9347187  1.0831983
           [,19]      [,20]      [,21]       [,22]      [,23]        [,24]
[1,] -0.02948362  1.1872102 0.54681574 -0.60784111  0.7215309 -0.329079736
[2,]  1.30841030 -0.3567140 0.49132117  1.07622314  1.0472770 -0.229793207
[3,]  1.97790329  1.2122385 0.05486097 -0.57642579  0.6316979  0.007676873
[4,] -2.31286660 -0.6939527 0.95968479  1.09862636 -0.4628356 -0.258654597
[5,]  0.12994804  1.3560616 0.17558627  1.40734169 -0.4986430  0.656339682
[6,] -2.50608663  0.9057313 1.19085425  0.03665615  1.8248174  0.086403496
           [,25]       [,26]      [,27]       [,28]       [,29]       [,30]
[1,]  0.06386908 -0.18264862 -0.9119685  0.71753201  1.05076285  0.89113145
[2,]  0.08015083 -1.39196798  0.6037755 -0.01430538 -0.07179733  0.91232151
[3,] -0.03688771  0.70824781  0.2138470 -0.14650171  0.11673662 -0.07131369
[4,]  0.13706006  0.18896582 -1.3806745 -0.06295297  0.97786651  1.13969595
[5,]  1.60949042 -0.21736688 -0.7994792 -0.35249734 -1.03154201 -0.51180167
[6,] -0.93868125  0.07163463  0.5170759 -0.02792490  2.19912933 -1.03759340
          [,31]       [,32]      [,33]      [,34]      [,35]      [,36]
[1,] -1.1963710  0.50989212 -0.2833076 -0.1109675 -0.1995761 -0.4819800
[2,]  0.1988017 -0.71536427 -1.7363750  0.2031390  0.7013954  1.6689437
[3,] -0.2142074 -0.40552154  0.6781351 -0.6757139  0.8477244  0.0199130
[4,]  0.1154809 -0.01252119  1.0620052  1.0741201  0.2288555 -0.2281483
[5,]  1.4809528  1.10088104  0.1499669  0.9673503 -0.1678426  0.2412909
[6,] -1.7794439  0.95737579  2.1944099  0.7700868 -0.6305805  0.3327065
          [,37]      [,38]        [,39]       [,40]     [,41]      [,42]
[1,] -0.2253496 -0.3264791 -1.688285063 -0.31538640 0.3075571 -1.0426910
[2,] -0.5265199  0.1353611 -1.075619561 -0.21114633 1.2673491 -0.1839484
[3,] -0.6424590  0.9596153 -0.396138667  0.40194336 1.4424067  0.1792667
```

```
[4,]  1.7167260  1.3431970  0.006904831  0.06096517 0.6023509 -0.6686468
[5,]  1.3234542  0.8142960  0.129844942 -0.08613139 0.4911252 -1.0043865
[6,]  0.1020486  0.5311642  1.684630296 -0.41713398 1.6725776  0.1367809
          [,43]       [,44]      [,45]       [,46]      [,47]      [,48]
[1,] -1.2346408  0.49186905  0.7409523 -0.86948531  0.4604544 -1.5678405
[2,]  0.6331305 -0.05041947  0.7083717 -0.05428413 -2.2889357  0.7410197
[3,]  1.0203876  1.24220506 -1.9020404 -0.50230446  1.2010044 -1.1983584
[4,] -0.3306417 -0.01164910 -1.0762147  0.22550387 -0.5422825 -1.5213027
[5,]  0.3751473 -1.50609464  0.1070659  0.97414676  0.7483055  1.5113067
[6,]  0.3470171 -0.93092817 -0.7851170 -0.97013833  0.6740285 -1.7291736
          [,49]      [,50]       [,51]
[1,] -0.3103677 -1.4389378  1.57953188
[2,] -1.3381571  0.5061491  0.04913839
[3,] -0.7136293 -1.2748504  0.84461037
[4,]  0.7427126  0.6399629  1.19023414
[5,]  0.5955325  0.4564364  2.42466530
[6,] -0.8537773 -0.7611737 -0.11393412
```

[59]:
```
lasso.reg = rlasso(Y ~ X, post = FALSE)
# use lasso, not-Post-lasso
# lasso.reg = rlasso(X, Y, post=FALSE)
sum.lasso <- summary(lasso.reg, all = FALSE)
# can also do print(lasso.reg, all=FALSE)
```

```
Call:
rlasso.formula(formula = Y ~ X, post = FALSE)

Post-Lasso Estimation:  FALSE

Total number of variables: 50
Number of selected variables: 4

Residuals:
     Min       1Q   Median       3Q      Max
-3.09472 -0.69599  0.08732  0.63980  3.42649

            Estimate
(Intercept)    0.097
1              4.775
2              4.922
3              4.622
16             0.003

Residual standard error: 1.02
Multiple R-squared:  0.9864
Adjusted R-squared:  0.9858
Joint significance test:
```

```
 the sup score statistic for joint significance test is 64.52 with a p-value of
0
```

[60]:
```
yhat.lasso = predict(lasso.reg)
#in-sample prediction
Xnew = matrix(rnorm(n * p), ncol = p)
# new X
Ynew = Xnew %*% beta + rnorm(n)
#new Y
yhat.lasso.new = predict(lasso.reg, newdata = Xnew)
#out-of-sample prediction
post.lasso.reg = rlasso(Y ~ X, post = TRUE)
#now use post-lasso
print(post.lasso.reg, all = FALSE)
```

```
Call:
rlasso.formula(formula = Y ~ X, post = TRUE)

(Intercept)           1           2           3
    0.06043     4.93977     5.09233     4.87495
```

[61]:
```
yhat.postlasso = predict(post.lasso.reg)
#in-sample prediction
yhat.postlasso.new = predict(post.lasso.reg, newdata = Xnew)
#out-of-sample prediction
MAE <- apply(cbind(abs(Ynew - yhat.lasso.new), abs(Ynew - yhat.postlasso.new)),
 ↪2,
mean)
names(MAE) <- c("lasso MAE", "Post-lasso MAE")
print(MAE, digits = 2)
```

```
     lasso MAE Post-lasso MAE
          0.82           0.74
```

5 Inference on Target Regression Coefficients

[62]:
```
#4.1 Intuition for the Orthogonality Principle in Linear Models via Partialling
 ↪Out
set.seed(1)
n = 5000
p = 20
X = matrix(rnorm(n * p), ncol = p)
colnames(X) = c("d", paste("x", 1:19, sep = ""))
xnames = colnames(X)[-1]
beta = rep(1, 20)
y = X %*% beta + rnorm(n)
```

```r
dat = data.frame(y = y, X)
r_data("02", dat)
```

```
           y          d         x1         x2         x3         x4         x5
1 -3.1877493 -0.6264538 -1.5163733 -0.8043316 -0.2139090  0.2353485  0.1965621
2  1.9275357  0.1836433  0.6291412 -1.0565257 -0.1067233  0.2448250 -0.4199427
3  0.9080673 -0.8356286 -1.6781940 -1.0353958 -0.4645893 -0.6421869  1.1632695
4  0.5918363  1.5952808  1.1797811 -1.1855604 -0.6842725 -1.9348085 -0.4057560
5  4.1376179  0.3295078  1.1176545 -0.5004395 -0.7908007  1.0386957  0.7440987
6 -6.4121254 -0.8204684 -1.2377359 -0.5249887 -0.3389638 -0.2835501  0.4766106
          x6         x7         x8         x9        x10         x11        x12
1  0.6179223  0.6986309 -0.2212571  0.5376559  0.5258908 -0.13104432  0.3413341
2  0.8935057 -1.1650711  0.3517935  1.2581705 -0.4875444 -1.42409627  0.4136665
3 -0.4277562  0.1713505  0.1606019 -0.6433912  1.1382508 -0.36966264  0.1220357
4 -0.2999012  1.1867162 -0.1240523  0.4578091  1.2151344  0.85504509 -1.5893806
5 -0.5319833  0.4107177  0.6598739  0.7720375 -0.4248307  0.97814426 -0.7874385
6  1.7059816 -1.2413395 -0.5038493  0.4730224 -1.4508403 -0.07130345 -1.5920640
         x13         x14         x15         x16         x17         x18
1 -1.206344 -1.00203611  0.35002823 -1.55915937 -1.09726565 -0.09504307
2  0.274206  0.02590761  0.40532916  0.20166217  2.42558030 -0.38805939
3  0.186822 -0.44814178  0.08469671  1.04017610  1.05186103  2.13657003
4 -0.263768  0.84323332  0.23416819  0.07195772 -0.08834213  0.55661945
5 -1.091930 -0.21846310 -0.15503122 -0.01526544  2.63979291 -0.59094164
6 -1.322318  0.47678629 -0.27158713  0.33938598 -1.72223733  1.52014345
         x19
1  0.9906532
2 -0.7241005
3  0.6700465
4 -0.5701159
5  0.7223508
6 -0.6084835
```

[63]:
```r
fmla = as.formula(paste("y ~ ", paste(colnames(X), collapse = "+")))
full.fit = lm(fmla, data = dat)
summary(full.fit)$coef["d", 1:2]
```

Estimate 0.978074548374755 **Std. Error** 0.0137122468163772

[64]:
```r
fmla.y = as.formula(paste("y ~ ", paste(xnames, collapse = "+")))
fmla.d = as.formula(paste("d ~ ", paste(xnames, collapse = "+")))
rY = lm(fmla.y, data = dat)$res
rD = lm(fmla.d, data = dat)$res
partial.fit.ls = lm(rY ~ rD)
summary(partial.fit.ls)$coef["rD", 1:2]
```

Estimate 0.978074548374761 **Std. Error** 0.0136861583043357

[65]:
```
rY = rlasso(fmla.y, data = dat)$res
rD = rlasso(fmla.d, data = dat)$res
partial.fit.postlasso = lm(rY ~ rD)
summary(partial.fit.postlasso)$coef["rD", 1:2]
```

Estimate	0.972738704584573	**Std. Error**	0.0136867656564768

6 Instrumental Variable Esimation in a High-Dimensional Setting

[66]:
```
Eff = rlassoEffect(X[, -1], y, X[, 1], method = "partialling out")
summary(Eff)$coef[, 1:2]
```

Estimate.	0.972738704584573	**Std. Error**	0.0136867656564768

[67]:
```
Eff = rlassoEffect(X[, -1], y, X[, 1], method = "double selection")
summary(Eff)$coef[, 1:2]
```

Estimate.	0.978074548374755	**Std. Error**	0.0141562427002124

[68]:
```
#4.2 Inference confidence Intervals and Significance Testing
set.seed(1)
n = 100
#sample size
p = 100
# number of variables
s = 3
# nubmer of non-zero variables
X = matrix(rnorm(n * p), ncol = p)
colnames(X) <- paste("X", 1:p, sep = "")
beta = c(rep(3, s), rep(0, p - s))
y = 1 + X %*% beta + rnorm(n)
data = data.frame(cbind(y, X))
colnames(data)[1] <- "y"
fm = paste("y ~", paste(colnames(X), collapse = "+"))
fm = as.formula(fm)

r_data(name = "03", data)
```

```
           y         X1          X2         X3         X4         X5
1 -2.3165875 -0.6264538 -0.62036668  0.4094018  0.8936737  1.0744410
2  5.6873718  0.1836433  0.04211587  1.6888733 -1.0472981  1.8956548
3 -0.5152813 -0.8356286 -0.91092165  1.5865884  1.9713374 -0.6029973
4  4.0816450  1.5952808  0.15802877 -0.3309078 -0.3836321 -0.3908678
5 -7.3313767  0.3295078 -0.65458464 -2.2852355  1.6541453 -0.4162220
6 10.8084527 -0.8204684  1.76728727  2.4976616  1.5122127 -0.3756574
          X6         X7          X8          X9        X10         X11
1  0.07730312 -0.3410670 -0.70756823 -1.08690882 -1.5414026  1.13496509
2 -0.29686864  1.5024245  1.97157201 -1.82608301  0.1943211  1.11193185
```

```
3 -1.18324224  0.5283077 -0.08999868  0.99528181  0.2644225 -0.87077763
4  0.01129269  0.5421914 -0.01401725 -0.01186178 -1.1187352  0.21073159
5  0.99160104 -0.1366734 -1.12345694 -0.59962839  0.6509530  0.06939565
6  1.59396745 -1.1367339 -1.34413012 -0.17794799 -1.0329002 -1.66264885
         X12        X13        X14        X15        X16         X17        X18
1  0.2418959 -1.5570357  0.3412484  1.5468813  0.8500435  0.34419403  1.6212029
2 -1.1327594  1.9231637  1.3161672  0.1789210 -0.9253130  0.01271984 -0.3291028
3  1.4899074 -1.8568296 -0.9597765 -0.2825466  0.8935812 -0.87345013 -2.3264095
4 -0.2482471 -2.1061184 -1.2055752 -0.7672988 -0.9410097  0.34280028  2.1929980
5  0.1835837  0.6976485  1.5675731 -0.5764042  0.5389521 -0.17738775 -1.0824800
6  0.4048710  0.9074444  0.2252858 -0.9148558 -0.1819744  0.92143325 -0.5063610
         X19         X20         X21         X22        X23        X24
1  0.7140855 -0.57099429 -0.88614959 -1.34105095  0.9169380 -0.3116892
2  0.5813846  0.28653902 -1.92225490 -0.04570723  0.8092731  0.2057491
3 -0.1467239  1.14761986  1.61970074  2.18799112 -0.7116223 -0.6539869
4  1.5069818  0.13955870  0.51926990  1.42209580 -2.6895852 -1.1532577
5 -0.2795326  0.08892661 -0.05584993  0.18324702 -0.5670470  0.5274909
6  2.0277387 -2.63015932  0.69641761 -0.65293284  1.2991988  1.3939191
         X25        X26         X27        X28         X29         X30
1 -0.3743289 -1.8054836  0.94033680 -0.4053392 -2.10406017  0.78104120
2  0.9953538 -0.6780407  0.78785519  1.9406715 -0.08443947 -0.04650931
3  0.1021435 -0.4733581  0.08694194  0.4849653  0.75632942  0.09576593
4  1.4829437  1.0274171  0.03280097 -0.2020973 -1.58071605 -1.33525787
5  0.5600487 -0.5973876  1.55285735 -1.1696286  0.70724595  0.54878591
6  0.1424510  1.1598494 -2.40487804 -0.3698461 -1.04598767 -1.90524352
         X31         X32        X33        X34         X35        X36
1  0.7391149  0.31570474 -0.1131544 -0.4456043 -1.04818566  0.5205997
2  0.3866087  0.16389139  0.8564798 -0.6763940 -0.42554881  0.3775619
3  1.2963972  0.95765836 -0.1855841  0.4116056 -0.23487313 -0.6236588
4 -0.8035584 -0.13109632  1.4280518 -0.5868514  1.19028909 -0.5726105
5 -1.6026257 -0.04676214  2.0817674 -1.2743676  0.54071372  0.3125012
6  0.9332510  1.08256389 -0.5674901 -0.8968956 -0.08926451 -0.7074278
         X37        X38        X39       X40        X41        X42         X43
1  1.7290728  0.8681650 -0.7280986 0.2618973 -1.1346302  0.4083129 -1.47983426
2 -0.8149923 -1.4843701 -0.2467004 0.1076991  0.7645571  0.4260585  1.02834657
3 -1.6902330 -0.4008993 -0.6136157 0.8309216  0.5707101 -1.1011658 -2.22105108
4  1.4909445 -0.6393477  0.1039478 0.8612745 -1.3516939 -0.3323497 -1.63855763
5  0.7036331  0.2163910 -0.8005786 0.3303093 -2.0298855  0.2302076  0.35723943
6 -0.9626936 -1.2611376  1.3237039 1.0899624  0.5904787 -1.1711534  0.02628372
          X44        X45        X46         X47         X48        X49
1 -0.40774766 -0.8396835  1.5579537 -0.62552459 -0.04479249 -1.1423104
2  0.99756662  0.4461303 -0.7292970 -1.42555255  0.56742135  1.1754607
3 -0.96926774 -0.3654167 -1.5039509  0.12344768  0.69540173 -0.4741456
4  0.75862510  0.5391799 -0.5667870 -0.57635310  1.24035018  1.6443085
5  0.08275072 -0.8085769 -2.1044536  0.09904507 -0.06951935 -0.3284536
6 -0.96894852 -0.4844113  0.5307319  1.12554493 -0.23488875  0.1529221
          X50        X51        X52        X53         X54         X55
1 -2.11988437 -1.5163733  0.8830833  0.4577942 -1.44364515  0.34927903
```

```
2 -0.66156344  0.6291412  1.3787128  1.6463162  0.68173681  0.95314451
3  0.31801209 -1.6781940  1.3423237  1.3117039  0.04923993 -2.22791423
4 -0.09837842  1.1797811 -0.7565422  0.1758473  0.07966677 -0.77626436
5 -0.03711058  1.1176545 -0.2990669 -0.7003017 -0.11879819 -0.07033167
6 -0.94716586 -1.2377359  1.0029136  0.8740681  0.73389984  1.26068980
         X56        X57         X58        X59         X60         X61
1 -1.1378698  0.6931984 -0.33228301 -0.1324555 -0.01338997 -0.61882708
2 -0.9518105 -0.1964654 -0.75639689 -0.3057390  0.89797626 -1.10942196
3  1.6192595  0.4277386  0.02701197  0.4773073 -0.85715025 -2.17033523
4  0.1678136  0.5584194 -0.21197870  0.2899630 -3.23261042 -0.03130307
5 -0.9081778 -0.8484908 -0.80742683 -0.4389847 -0.14938681 -0.26039848
6  1.3417959  0.3932101  1.32708947 -0.7283198 -0.36177371  0.53443047
          X62        X63         X64         X65        X66        X67
1  1.16866564 -1.2257122  0.06158714  0.43061750  0.8871888 -0.2903608
2  2.02175508  0.1433413  0.56738070 -0.20312359 -0.3776280  0.9465863
3 -0.26573855 -0.5297313  0.53243870  0.92401507  0.1140808 -0.9679713
4 -0.07937292  0.3376394 -0.94140665 -1.36990446 -1.7242394 -1.2862308
5 -0.26026605  0.1096895  0.46536548 -1.50960376  0.7436886  1.3106510
6  0.65082651  1.5117803  1.10889037  0.08513647 -0.8857317 -0.9900365
         X68        X69        X70        X71        X72         X73        X74
1  0.8667786  0.1500674  0.8171146 -1.3254177 -0.2084137  0.07489798  1.2350840
2  0.2502538  0.6000764  0.1818066  0.9519797 -2.0155207  0.45019776  1.7002581
3 -0.7117007 -0.9283509  1.2532384  0.8600044 -1.0348842 -0.16056416 -0.4484478
4  0.5852223 -0.3040067  0.6069871  1.0607903 -0.2030673  0.84669312 -0.4264031
5  1.6453477 -0.7667067  0.2825354 -0.3505840  0.2044387 -0.02494173  0.3381126
6 -0.0509118 -1.7372357 -0.3089424 -0.1307656 -0.2584211  0.40911405  0.2927457
         X75        X76         X77        X78         X79        X80
1  0.6917708 -0.7948034  1.19622325 -0.9067851  0.56151328 -1.2274212
2 -1.4548485  0.6060846 -0.03550621 -0.8600897 -0.37612365  0.3579550
3  0.1238508 -1.0624674  0.80889396  1.5293101 -0.73666064  1.3680534
4 -1.6157229  1.0192005 -0.86704016 -0.3291295  0.05441946  1.5579761
5  0.8594476  0.1776102  0.01576747 -0.6553643 -0.62641811 -1.3145953
6  1.4352811 -1.0309747  0.33461765 -1.2012061  0.19662709 -0.5494434
         X81         X82        X83         X84         X85         X86
1  0.2637034 -0.47118610  0.2607109 -0.59298107  1.67555392  0.99010104
2 -0.8294518 -0.14928304 -0.9985321 -2.24818494 -0.23087887 -0.06643542
3 -1.4616348  0.51092425  3.3769118  0.53802260 -0.01518735  0.25797379
4  1.6839902  0.64200743  0.3995816 -1.61412479  0.28768115 -0.62145348
5 -1.5443243  0.02136446  2.3400609  0.01546701 -0.22471407 -0.77645263
6 -0.1908871 -0.58611920  2.0036462 -0.27797898 -0.39701376  0.31534275
         X87       X88        X89        X90        X91         X92         X93
1 -1.3469730 0.1002483  0.9636739 -1.1225773 -1.2171201  1.15394424 -0.72512090
2  0.5225039 0.3346909 -1.5261577  0.1203985 -0.9462293  1.31344739 -0.05514386
3  1.0039196 0.2789446  1.2213787 -0.2551244  0.0914098 -1.56846751  2.95813695
4  0.7588889 0.5040394 -0.3196803 -0.9364678  0.7013513  0.08680971 -0.38880448
5  0.4777187 1.5681529  0.5658068 -0.8101785  0.6734224 -0.03771153 -0.72195939
6 -0.4667302 0.2133896 -0.9163541 -0.6742103  1.2655534  0.52377161 -2.73897535
         X94         X95        X96        X97        X98        X99       X100
```

```
1 -0.5839239  1.10817150  1.2980378  1.0975878  0.7293022 -1.0096854  0.6183543
2  0.4908514  0.20770442 -1.4276760 -0.8995776 -1.0823819 -1.1837743  0.5798023
3 -0.6046265 -0.77437348  0.2427872 -1.1429702 -0.2678592 -1.1876126 -0.6658887
4 -0.3568455 -0.08472013 -0.2107006  0.7668381 -0.4968127  1.0732680  1.2307256
5 -0.4658754 -1.26100902  0.0801386 -1.0705925 -1.5365873 -0.9617255  0.9418813
6  1.8941562  0.10847615  1.5460849 -0.8389841  0.2739188  0.9681514  1.4503185
```

[69]:
```
lasso.effect = rlassoEffects(fm, I = ~X1 + X2 + X3 + X50, data = data)
print(lasso.effect)
```

```
Call:
rlassoEffects.formula(formula = fm, data = data, I = ~X1 + X2 +
    X3 + X50)

Coefficients:
     X1       X2       X3      X50
2.94448  3.04127  2.97540  0.07196
```

[70]:
```
summary(lasso.effect)
```

```
[1] "Estimates and significance testing of the effect of target variables"
    Estimate. Std. Error t value Pr(>|t|)
X1    2.94448    0.08815  33.404   <2e-16 ***
X2    3.04127    0.08389  36.253   <2e-16 ***
X3    2.97540    0.07804  38.127   <2e-16 ***
X50   0.07196    0.07765   0.927    0.354
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[71]:
```
confint(lasso.effect)
```

A matrix: 4 × 2 of type dbl

	2.5 %	97.5 %
X1	2.77171308	3.1172421
X2	2.87685121	3.2056979
X3	2.82244962	3.1283583
X50	-0.08022708	0.2241377

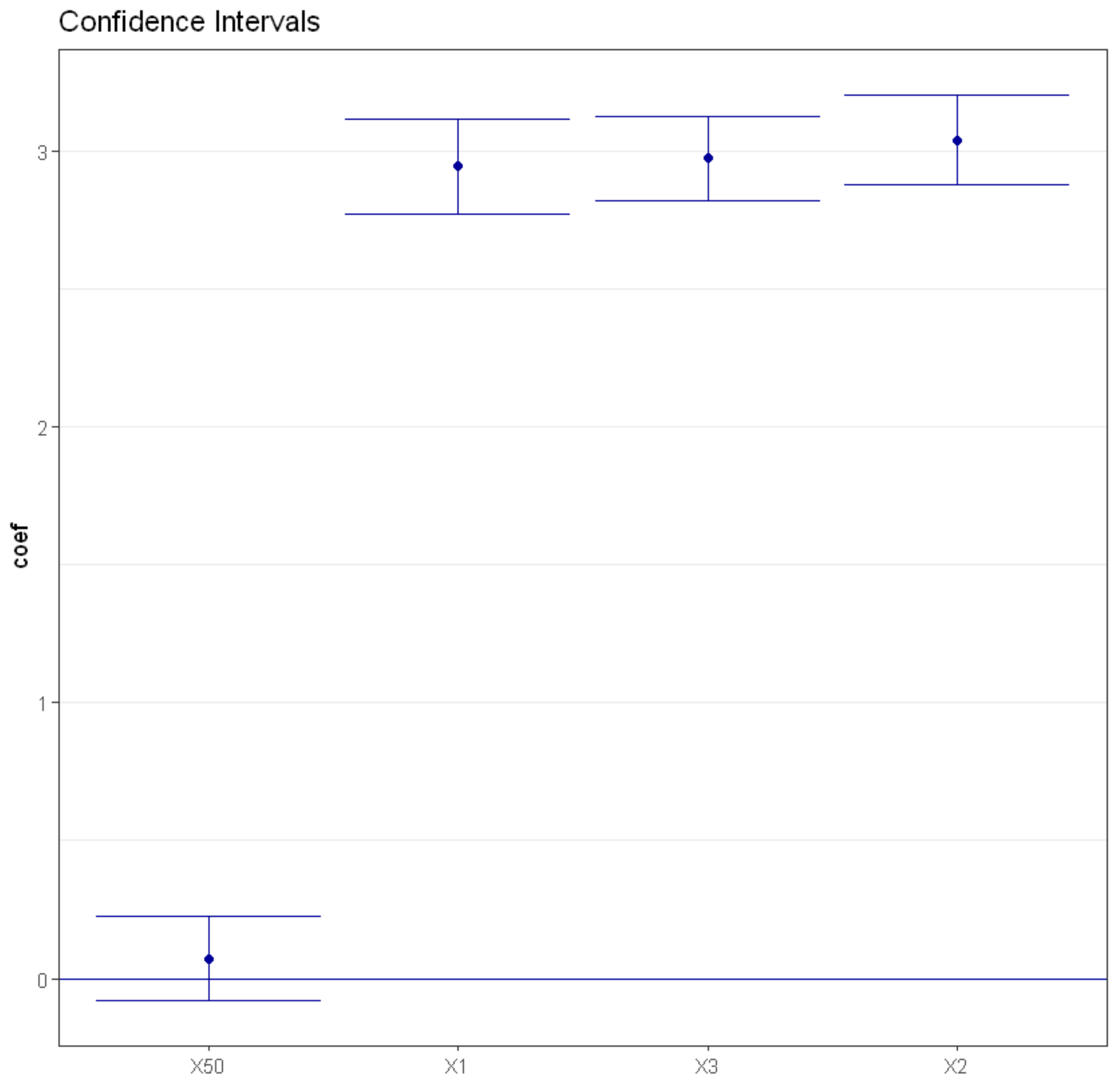
[72]:
```
confint(lasso.effect, level = 0.95, joint = TRUE)
```

A matrix: 4 × 2 of type dbl

	2.5 %	97.5 %
X1	2.7279477	3.1610075
X2	2.8371214	3.2454278
X3	2.7833176	3.1674903
X50	-0.1154509	0.2593615

[73]:
```
plot(lasso.effect, main = "Confidence Intervals")
```

```
Warning message:
"Ignoring unknown aesthetics: width, h"
```



[74]:
```
cps2012$cps2012
```

```
NULL
```

[75]:
```
# # 4.3

library(hdm)
cps2012 = head(hdm::cps2012, 200)
# cps2012 = data(cps2012)
```

```
X <- model.matrix(~-1 + female + female:(widowed + divorced + separated +␣
 ↪nevermarried +
hsd08 + hsd911 + hsg + cg + ad + mw + so + we + exp1 + exp2 + exp3) + +(widowed␣
 ↪+
divorced + separated + nevermarried + hsd08 + hsd911 + hsg + cg + ad + mw + so +
we + exp1 + exp2 + exp3)^2, data = cps2012)
# dim(X)
# [1] 29217
# 136
X <- X[, which(apply(X, 2, var) != 0)]
# exclude all constant variables
# dim(X)
# [1] 29217
# 116
index.gender <- grep("female", colnames(X))
y <- cps2012$lnw
index.gender
```

1. 1 2. 13 3. 14 4. 15 5. 16 6. 17 7. 18 8. 19 9. 20 10. 21

[76]:
```
Sys.sleep(10)
effects.female <- rlassoEffects(x = X, y = y, index = index.gender)
summary(effects.female)
```

```
[1] "Estimates and significance testing of the effect of target variables"
                     Estimate. Std. Error t value Pr(>|t|)
female              -3.619e-01  2.428e-01  -1.491    0.136
female:divorced      3.297e-01  2.375e-01   1.388    0.165
female:nevermarried  3.679e-01  2.271e-01   1.620    0.105
female:hsd911        7.180e+12  4.961e+12   1.447    0.148
female:hsg           8.463e-02  1.986e-01   0.426    0.670
female:cg            1.657e-01  2.005e-01   0.826    0.409
female:ad            2.722e-01  2.503e-01   1.087    0.277
female:exp1          3.149e-02  3.142e-02   1.002    0.316
female:exp2          3.526e-01  3.586e-01   0.983    0.325
female:exp3         -2.635e-03  2.775e-02  -0.095    0.924
```

[77]:
```
joint.CI <- confint(effects.female, level = 0.95, joint = TRUE)
joint.CI
```

A matrix: 10 × 2 of type dbl

	2.5 %	97.5 %
female	-2.409128e+00	1.685259e+00
female:divorced	-1.065406e+00	1.724708e+00
female:nevermarried	-1.127857e+00	1.863675e+00
female:hsd911	-3.117441e+13	4.553491e+13
female:hsg	-1.777279e+00	1.946531e+00
female:cg	-1.529498e+00	1.860816e+00
female:ad	-8.197200e-01	1.364067e+00
female:exp1	-1.684121e-01	2.313928e-01
female:exp2	-2.385428e+00	3.090589e+00
female:exp3	-2.212582e-01	2.159888e-01

[78]:
```
Sys.sleep(7)
effects.female <- rlassoEffects(lnw ~ female + female:(widowed + divorced +␣
 ↪separated +
nevermarried + hsd08 + hsd911 + hsg + cg + ad + mw + so + we + exp1 + exp2 +
exp3) + (widowed + divorced + separated + nevermarried + hsd08 + hsd911 + hsg +
cg + ad + mw + so + we + exp1 + exp2 + exp3)^2, data = cps2012, I = ~female +
female:(widowed + divorced + separated + nevermarried + hsd08 + hsd911 + hsg +
cg + ad + mw + so + we + exp1 + exp2 + exp3))
```

[79]:
```
# 4.4

GrowthData = head(hdm::GrowthData, 200)
# data(GrowthData)
# dim(GrowthData)
# [1] 90 63
y = GrowthData[, 1, drop = F]
d = GrowthData[, 3, drop = F]
X = as.matrix(GrowthData)[, -c(1, 2, 3)]
varnames = colnames(GrowthData)
```

[80]:
```
xnames = varnames[-c(1, 2, 3)]
# names of X variables
dandxnames = varnames[-c(1, 2)]
# names of D and X variables
# create formulas by pasting names (this saves typing times)
fmla = as.formula(paste("Outcome ~ ", paste(dandxnames, collapse = "+")))
ls.effect = lm(fmla, data = GrowthData)
```

[81]:
```
dX = as.matrix(cbind(d, X))
lasso.effect = rlassoEffect(x = X, y = y, d = d, method = "partialling out")
summary(lasso.effect)
```

```
[1] "Estimates and significance testing of the effect of target variables"
     Estimate. Std. Error t value Pr(>|t|)
[1,]  -0.04981    0.01394  -3.574 0.000351 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[82]:
```
dX = as.matrix(cbind(d, X))
doublesel.effect = rlassoEffect(x = X, y = y, d = d, method = "double␣
↪selection")
summary(doublesel.effect)
```

```
[1] "Estimates and significance testing of the effect of target variables"
         Estimate. Std. Error t value Pr(>|t|)
gdpsh465  -0.05001    0.01579  -3.167  0.00154 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[83]:
```
library(xtable)
table = rbind(summary(ls.effect)$coef["gdpsh465", 1:2], summary(lasso.
↪effect)$coef[,
1:2], summary(doublesel.effect)$coef[, 1:2])
colnames(table) = c("Estimate", "Std. Error")
#names(summary(full.fit)£coef)[1:2]
rownames(table) = c("full reg via ols", "partial reg
via post-lasso ", "partial reg via double selection")
tab = xtable(table, digits = c(2, 2, 5))
tab
```

A xtable: 3 × 2

	Estimate <dbl>	Std. Error <dbl>
full reg via ols	-0.009377989	0.02988773
partial reg via post-lasso	-0.049811465	0.01393636
partial reg via double selection	-0.050005855	0.01579138

7 Inference on Treatment Effects in a Hight-Dimensional Setting

[84]:
```
#5.1
AJR = head(hdm::AJR, 200)
y = AJR$GDP
d = AJR$Exprop
z = AJR$logMort
x = model.matrix(~-1 + (Latitude + Latitude2 + Africa + Asia + Namer + Samer)^2,
data = AJR)
# dim(AJR)
```

[85]:
```
AJR.Xselect = rlassoIV(GDP ~ Exprop + (Latitude + Latitude2 + Africa + Asia +␣
↪Namer +
Samer)^2 | logMort + (Latitude + Latitude2 + Africa + Asia + Namer + Samer)^2,
```

```
data = AJR, select.X = TRUE, select.Z = FALSE)
summary(AJR.Xselect)
```

```
[1] "Estimation and significance testing of the effect of target variables in
the IV regression model"
       coeff.    se. t-value p-value
Exprop 0.8450 0.2699   3.131 0.00174 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[86]:

```
confint(AJR.Xselect)
```

```
           2.5 %   97.5 %
Exprop 0.3159812 1.374072
```

[87]:

```
fmla.y = GDP ~ (Latitude + Latitude2 + Africa + Asia + Namer + Samer)^2
fmla.d = Exprop ~ (Latitude + Latitude2 + Africa + Asia + Namer + Samer)^2
fmla.z = logMort ~ (Latitude + Latitude2 + Africa + Asia + Namer + Samer)^2
rY = lm(fmla.y, data = AJR)$res
rD = lm(fmla.d, data = AJR)$res
rZ = lm(fmla.z, data = AJR)$res
# ivfit.lm = tsls(y=rY,d=rD, x=NULL, z=rZ, intercept=FALSE)
ivfit.lm = tsls(rY ~ rD | rZ, intercept = FALSE)
print(cbind(ivfit.lm$coef, ivfit.lm$se), digits = 3)
```

```
   [,1] [,2]
rD 1.27 1.73
```

[88]:

```
rY = rlasso(fmla.y, data = AJR)$res
rD = rlasso(fmla.d, data = AJR)$res
rZ = rlasso(fmla.z, data = AJR)$res
# ivfit.lasso = tsls(y=rY,d=rD, x=NULL, z=rZ, intercept=FALSE)
ivfit.lasso = tsls(rY ~ rD | rZ, intercept = FALSE)
print(cbind(ivfit.lasso$coef, ivfit.lasso$se), digits = 3)
```

```
    [,1] [,2]
rD 0.845 0.27
```

[89]:

```
EminentDomain = head(hdm::EminentDomain, 200)
z <- as.matrix(EminentDomain$logGDP$z)
x <- as.matrix(EminentDomain$logGDP$x)
y <- EminentDomain$logGDP$y
d <- EminentDomain$logGDP$d
x <- x[, apply(x, 2, mean, na.rm = TRUE) > 0.05]
#
z <- z[, apply(z, 2, mean, na.rm = TRUE) > 0.05]
#
```

[90]:
```
ED.ols = lm(y ~ cbind(d, x))
ED.2sls = tsls(y = y, d = d, x = x, z = z[, 1:2], intercept = FALSE)
```

[91]:
```
lasso.IV.Z = rlassoIV(x = x, d = d, y = y, z = z, select.X = FALSE, select.Z =
  ↪TRUE)
# or lasso.IV.Z = rlassoIVselectZt(x=X, d=d, y=y, z=z)
summary(lasso.IV.Z)
```

```
[1] "Estimates and significance testing of the effect of target variables in the
IV regression model"
   coeff.    se. t-value p-value
d1 0.4146 0.2902   1.428   0.153
```

[92]:
```
confint(lasso.IV.Z)
```

```
        2.5 %    97.5 %
d1 -0.1542764 0.9834796
```

[93]:
```
lasso.IV.XZ = rlassoIV(x = x, d = d, y = y, z = z, select.X = TRUE, select.Z =
  ↪TRUE)
summary(lasso.IV.XZ)
```

```
Estimates and Significance Testing of the effect of target variables in the IV
regression model
     coeff.      se. t-value p-value
d1 -0.02383  0.12851  -0.185   0.853
```

[94]:
```
confint(lasso.IV.XZ)
```

```
        2.5 %    97.5 %
d1 -0.2757029 0.2280335
```

[95]:
```
library(xtable)
table = matrix(0, 4, 2)
table[1, ] = summary(ED.ols)$coef[2, 1:2]
table[2, ] = cbind(ED.2sls$coef[1], ED.2sls$se[1])
table[3, ] = summary(lasso.IV.Z)[, 1:2]
```

```
[1] "Estimates and significance testing of the effect of target variables in the
IV regression model"
   coeff.    se. t-value p-value
d1 0.4146 0.2902   1.428   0.153
```

[96]:
```
table[4, ] = summary(lasso.IV.XZ)[, 1:2]
```

```
Estimates and Significance Testing of the effect of target variables in the IV
regression model
     coeff.      se. t-value p-value
d1 -0.02383  0.12851  -0.185   0.853
```

```
[97]: colnames(table) = c("Estimate", "Std. Error")
rownames(table) = c("ols regression", "IV estimation ", "selection on Z",
 ↪"selection on X and Z")
tab = xtable(table, digits = c(2, 2, 7))
tab
```

A xtable: 4 × 2

	Estimate <dbl>	Std. Error <dbl>
ols regression	0.007864732	0.009865927
IV estimation	-0.010733269	0.033766362
selection on Z	0.414601641	0.290249208
selection on X and Z	-0.023834697	0.128506538

```
[98]: pension = head(hdm::pension, 200)
y = pension$tw
d = pension$p401
z = pension$e401
X = pension[, c("i2", "i3", "i4", "i5", "i6", "i7", "a2", "a3", "a4", "a5",
 ↪"fsize",
"hs", "smcol", "col", "marr", "twoearn", "db", "pira", "hown")]
# simple model
xvar = c("i2", "i3", "i4", "i5", "i6", "i7", "a2", "a3", "a4", "a5", "fsize",
 ↪"hs",
"smcol", "col", "marr", "twoearn", "db", "pira", "hown")
xpart = paste(xvar, collapse = "+")
form = as.formula(paste("tw ~ ", paste(c("p401", xvar), collapse = "+"), "|",
 ↪paste(xvar,
collapse = "+")))
formZ = as.formula(paste("tw ~ ", paste(c("p401", xvar), collapse = "+"), "|",
 ↪paste(c("e401",
xvar), collapse = "+")))
```

```
[99]: pension.ate = rlassoATE(form, data = pension)
summary(pension.ate)
```

```
Error in check_binary(d): Treatment variable and Instrumental Variable should be
 ↪binary (0/1)!
Traceback:

1. rlassoATE(form, data = pension)
2. rlassoATE.formula(form, data = pension)
```

```
3. check_binary(d)
4. stop("Treatment variable and Instrumental Variable should be binary (0/1)!")
```

[ ]:
```
pension.atet = rlassoATET(form, data = pension)
summary(pension.atet)
```

```
Estimation and significance testing of the treatment effect
Type: ATET
Bootstrap: not applicable
   coeff.   se. t-value p-value
TE  12628  2944   4.289 1.8e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

7.1 Error

[ ]:
```
# pension.late = rlassoLATE(X, d, y, z)
# pension.late = rlassoLATE(formZ, data=pension)
# summary(pension.late)
```

[ ]:
```
# pension.latet = rlassoLATET(X, d, y, z)
```

[ ]:
```
xvar2 <- paste("(", xvar, ")^2", sep = "")
formExt = as.formula(paste("tw ~ ", paste(c("p401", xvar2), collapse = "+"),␣
  ↪"|",
paste(xvar2, collapse = "+")))
formZExt = as.formula(paste("tw ~ ", paste(c("p401", xvar2), collapse = "+"),␣
  ↪"|",
paste(c("e401", xvar2), collapse = "+")))
```

[ ]:
```
pension.ate = rlassoATE(X, z, y)
pension.atet = rlassoATET(X, z, y)
# pension.late = rlassoLATE(X, d, y, z)
# pension.latet = rlassoLATET(X, d, y, z)
```

8 The Lasso Methods for Discovery of Significant Causes amongst Many Potential Causes, with Many Controls

[ ]:
```
set.seed(1)
n = 100
p1 = 20
p2 = 20
D = matrix(rnorm(n * p1), n, p1)
# Causes
```

```
W = matrix(rnorm(n * p2), n, p2)
X = cbind(D, W)
# Regressors
Y = D[, 1] * 5 + W[, 1] * 5 + rnorm(n)
#Outcome
confint(rlassoEffects(X, Y, index = c(1:p1)), joint = TRUE)
```

A matrix: 20 × 2 of type dbl

	2.5 %	97.5 %
V1	4.5145877	5.21430498
V2	-0.3142909	0.30494650
V3	-0.3524109	0.18678880
V4	-0.2542430	0.28738914
V5	-0.2765802	0.27627177
V6	-0.3214676	0.29422684
V7	-0.2262507	0.30094168
V8	-0.0473541	0.47366372
V9	-0.1865636	0.39023520
V10	-0.2372356	0.26411185
V11	-0.3147091	0.20945872
V12	-0.3091905	0.26572176
V13	-0.1741550	0.37682465
V14	-0.3235734	0.38543162
V15	-0.3219763	0.31312486
V16	-0.2649505	0.33100700
V17	-0.1792080	0.41696169
V18	-0.3693247	0.04695928
V19	-0.1073109	0.39368776
V20	-0.2157182	0.25543839