Challenge (Regularization)

1 Introduction

I use 52 variables in GSS to find which of them has relatively more influence on people's attutude on whether a woman should continue to be able to have an abortion legally or not (abfelegl).

Below are the variables I choose:

CodeintheDataset	VarName
wrkstat	Last week were you working full time, part time, going to school, keeping house, or what?
occ10	RESPONDENT'S OCCUPATION
indus10	RESPONDENT'S INDUSTRY
marital	Are you currently–married, widowed, divorced, separated, or have you never been married?
paocc10	FATHER'S OCCUPATION
papres10	PRESTIGE OF RESPONDENT'S FATHER'S OCCUPATION
paind10	FATHER'S INDUSTRY
maocc10	MOTHER'S OCCUPATION
mapres10	PRESTIGE OF RESPONDENT'S MOTHER'S OCCUPATION
maind10	MOTHER'S INDUSTRY
childs	How many children have you ever had?
age	RESPONDENT'S AGE
res16	Which of the categories on this card comes closest to the type of place you were living in when you were 16 years old?
reg16	In what state or foreign country were you living when you were 16 years old?
family16	Were you living with both your own mother and father around the time you were 16?

CodeintheDataset	VarName
agekdbrn	How old were you when your first child was born?
educ	RESPONDENT'S EDUCATION
paeduc	RESPONDENT'S FATHER'S (FATHER SUBSTITUTE'S) EDUCATION
maeduc	RESPONDENT'S MOTHER'S (SUBSTITUTE MOTHER'S) EDUCATION
degree	RESPONDENT'S DEGREE
race	What race do you consider yourself?
mawrkgrw	Did your mother ever work for pay for as long as a year, while you were growing up?
incom16	Thinking about the time when you were 16 years old, compared with American families in general then, would you say your family income was far below average, below average, average, above average, or far above average?
income	In which of these groups did your total family income, from all sources, fall last year before taxes, that is?
region	REGION OF INTERVIEW
partyid	Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?
polviews	We hear a lot of talk these days about liberals and conservatives. I'm going to show you a seven-point scale on which the political views that people might hold are arranged from extremely liberal –point 1–to extremely conservative–point 7. Where would you place yourself on this scale?
postlife	Do you believe there is a life after death?
pray	About how often do you pray?
happy	Taken all together, how would you say things are these days –would you say that you are very happy, pretty happy, or not too happy?
conmedic	As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them? Medicine
class	If you were asked to use one of four names for your social class, which would you say you belong in: the lower class, the working class, the middle class, or the upper class?
rank	In our society there are groups which tend to be towards the top and those that are towards the bottom. Where would you put yourself on this scale?

CodeintheDataset	VarName
finrela	Compared with American families in general, would you say your family income is far below average, below average, average, above average, or far above average?
kidssol	When your children are at the age you are now, do you think their standard of living will be much better, somewhat better, about the same, somewhat worse, or much worse than yours is now?
sexeduc	Would you be for or against sex education in the public schools?
spanking	Do you strongly agree, agree, disagree, or strongly disagree that it is sometimes necessary to discipline a child with a good, hard spanking?
fechld	A working mother can establish just as warm and secure a relationship with her children as a mother who does not work. （whether you strongly agree, agree, disagree, or strongly disagree with it.）
fefam	It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family. （whether you strongly agree, agree, disagree, or strongly disagree with it.）
relpersn	To what extent do you consider yourself a religious person? Are you . . .
sprtprsn	To what extent do you consider yourself a spiritual person? Are you . . .
learnnew	My job requires that I keep learning new things （whether you strongly agree, agree, disagree, or strongly disagree with each of these statements.）
overwork	I have too much work to do everything well （whether you strongly agree, agree, disagree, or strongly disagree with each of these statements.）
promteok	The chances for promotion are good（whether you strongly agree, agree, disagree, or strongly disagree with each of these statements.）
rincblls	Do you feel that the income from your job alone is enough to meet your family's usual monthly expenses and bills?
hlthdays	During the past 30 days, for about how many days did your poor physical or mental health keep you from doing your usual activities, such as self-care, work, or recreation?
depress	Has a doctor, nurse, or other health professional EVER told you that you had Depression?
workfor1	Do you work for a for-profit company, a non-profit or not-for-profit organization, or for the government or a government agency?

CodeintheDataset	VarName
partners	How many sex partners have you had in the last 12 months?
evcrack	Have you ever, even once, used “crack”cocaine in chunk or rock form?
eth1	FIRST ETHNICITY MENTIONED
wordsum	Total number of correct words.(how people go about guessing words they do not know)

2 Preparation

First I conduct data management before formally analyzing them. Since the data of GSS comes from in ".dta" format, I choose to use stata to do preparation and further analysis. The code in stata is as follows:

```
clear all
version 14
set more off
cd "/Users/Documents/MicroEconometrics"

use "/Users/Documents/MicroEconometrics/GSS2018.dta", replace

keep  wrkstat occ10 indus10 ///
      marital paocc10 papres10 ///
      paind10 maocc10 mapres10 ///
      maind10 childs age agekdbrn ///
      educ paeduc maeduc degree race ///
      res16 reg16 family16 mawrkgrw ///
      incom16 income region partyid ///
      polviews postlife pray happy ///
      conmedic class rank finrela ///
      kidssol sexeduc spanking fechld ///
      fefam relpersn sprtprsn learnnew ///
      overwork promteok rincblls hlthdays ///
      depress workfor1 partners evcrack ///
      eth1 abfelegl wordsum

foreach checkvar of varlist wrkstat occ10 indus10 {
        replace ‘checkvar’ = . if ‘checkvar’ == .i
```

```
        replace `checkvar' = . if `checkvar' == .n
        replace `checkvar' = . if `checkvar' == .d
}
foreach checkvar of varlist marital paocc10 papres10 {
        replace `checkvar' = . if `checkvar' == .i
        replace `checkvar' = . if `checkvar' == .n
        replace `checkvar' = . if `checkvar' == .d
}
foreach checkvar of varlist paind10 maocc10 mapres10 {
        replace `checkvar' = . if `checkvar' == .i
        replace `checkvar' = . if `checkvar' == .n
        replace `checkvar' = . if `checkvar' == .d
}
foreach checkvar of varlist maind10 childs age agekdbrn {
        replace `checkvar' = . if `checkvar' == .i
        replace `checkvar' = . if `checkvar' == .n
        replace `checkvar' = . if `checkvar' == .d
}
foreach checkvar of varlist educ paeduc maeduc degree race {
        replace `checkvar' = . if `checkvar' == .i
        replace `checkvar' = . if `checkvar' == .n
        replace `checkvar' = . if `checkvar' == .d
}
foreach checkvar of varlist res16 reg16 family16 mawrkgrw {
        replace `checkvar' = . if `checkvar' == .i
        replace `checkvar' = . if `checkvar' == .n
        replace `checkvar' = . if `checkvar' == .d
}
foreach checkvar of varlist incom16 income region partyid {
        replace `checkvar' = . if `checkvar' == .i
        replace `checkvar' = . if `checkvar' == .n
        replace `checkvar' = . if `checkvar' == .d
}
foreach checkvar of varlist polviews postlife pray happy {
        replace `checkvar' = . if `checkvar' == .i
        replace `checkvar' = . if `checkvar' == .n
        replace `checkvar' = . if `checkvar' == .d
}
foreach checkvar of varlist conmedic class rank finrela {
        replace `checkvar' = . if `checkvar' == .i
```

```
        replace ‘checkvar’ = . if ‘checkvar’ == .n
        replace ‘checkvar’ = . if ‘checkvar’ == .d
}
foreach checkvar of varlist kidssol sexeduc spanking fechld {
        replace ‘checkvar’ = . if ‘checkvar’ == .i
        replace ‘checkvar’ = . if ‘checkvar’ == .n
        replace ‘checkvar’ = . if ‘checkvar’ == .d
}
foreach checkvar of varlist fefam relpersn sprtprsn learnnew{
        replace ‘checkvar’ = . if ‘checkvar’ == .i
        replace ‘checkvar’ = . if ‘checkvar’ == .n
        replace ‘checkvar’ = . if ‘checkvar’ == .d
}
foreach checkvar of varlist overwork promteok rincblls hlthdays{
        replace ‘checkvar’ = . if ‘checkvar’ == .i
        replace ‘checkvar’ = . if ‘checkvar’ == .n
        replace ‘checkvar’ = . if ‘checkvar’ == .d
}
foreach checkvar of varlist depress workfor1 partners evcrack {
        replace ‘checkvar’ = . if ‘checkvar’ == .i
        replace ‘checkvar’ = . if ‘checkvar’ == .n
        replace ‘checkvar’ = . if ‘checkvar’ == .d
}
foreach checkvar of varlist eth1 abfelegl wordsum {
        replace ‘checkvar’ = . if ‘checkvar’ == .i
        replace ‘checkvar’ = . if ‘checkvar’ == .n
        replace ‘checkvar’ = . if ‘checkvar’ == .d
}
```

3 Lasso

Then, I use 5 fold cross validation to find the optimal lambda of lasso, then get the regression results corresponding with this optimal lambda. The code in stata is as follows:

```
ssc install lassopack, replace
cvlasso   abfelegl wrkstat occ10 indus10 ///
          marital paocc10 papres10 ///
```

```
        paind10 maocc10 mapres10 ///
        maind10 childs age agekdbrn ///
        educ paeduc maeduc degree race ///
        res16 reg16 family16 mawrkgrw ///
        incom16 income region partyid ///
        polviews postlife pray happy ///
        conmedic class rank finrela ///
        kidssol sexeduc spanking fechld ///
        fefam relpersn sprtprsn learnnew ///
        overwork rincblls hlthdays ///
        depress evcrack eth1 wordsum, ///
        long nfolds(5) alpha(1.0)
```

Here is the result:

a	Lambda	MSPE	st. dev.
1	33.529703	0.94100638	0.05322184
2	**30.551016**	**0.93948975**	**0.05690603***
3	27.836948	0.94650811	0.06054795
4	25.36399	0.94880688	0.06196621
5	23.110723	0.95076051	0.06213154
6	21.05763	0.95968608	0.06184343
7	19.186929	0.9715781	0.062042
8	17.482415	0.98937165	0.06244187
9	15.929325	1.0130065	0.06541887
10	14.514208	1.042428	0.07064219
11	13.224806	1.0760984	0.07907597
12	12.04995	1.1080402	0.08916383
13	10.979466	1.141929	0.09753778
14	10.00408	1.1867062	0.10447866
15	9.1153456	1.2475813	0.11269912
16	8.3055634	1.3072832	0.12284751
17	7.56772	1.3719659	0.13401591
18	6.8954246	1.4400587	0.15458219
19	6.2828541	1.4982495	0.19569752
20	5.7247027	1.5923691	0.20553754

From this table, we get the optimal lambda and the corresponding regression with this following code:

```
cvlasso, lopt
```

And the results are listed as follows:

Selected	Lasso	Post-est OLS
happy	-0.0022546	-0.5008649
promteok	-0.0288106	-0.3056890
Partialled-out*		
_cons	2.1088930	3.6204113

4 Elastic Net

Then I conduct corss-validation elastic net with this dataset with several alpha, the code and results are as follows:

4.1 Elastic Net with $\alpha = 0.2$

```
cvlasso abfelegl wrkstat occ10 indus10 ///
marital paocc10 papres10 ///
paind10 maocc10 mapres10 ///
maind10 childs age agekdbrn ///
educ paeduc maeduc degree race ///
res16 reg16 family16 mawrkgrw ///
incom16 income region partyid ///
polviews postlife pray happy ///
conmedic class rank finrela ///
kidssol sexeduc spanking fechld ///
fefam relpersn sprtprsn learnnew ///
overwork promteok rincblls hlthdays ///
depress workfor1 partners evcrack ///
eth1 wordsum, long nfolds(5) alpha(0.2)
```

The code above yields the results as:

Selected	Elastic net (alpha=0.200)	Post-est OLS
papres10	0.0010458	-0.0009047
childs	0.0680969	0.0556009
agekdbrn	-0.0058746	-0.013785
partyid	0.0210815	0.0724655
polviews	0.0484773	0.0739597
postlife	-0.0636745	0.0431401
pray	-0.056648	-0.057334
happy	-0.3281692	-0.5990131
conmedic	0.0656661	0.1480275
kidssol	-0.1056309	-0.1676854
spanking	-0.0241453	-0.0578842
fechld	0.0757882	0.2629612
fefam	-0.1073232	0.0481283
overwork	-0.0147769	-0.0703576
promteok	-0.1098092	-0.1562279
hlthdays	0.016018	0.0200737
workfor1	0.1040439	0.197243
partners	-0.0512729	-0.2443124
eth1	0.0023778	0.006143
Partialled-out*		
_cons	2.9355073	2.7579731

4.2 Elastic Net with $\alpha = 0.4$

Changing the code above from $alpha(0.2)$ to $alpha(0.4)$ yields the results as:

Selected	Elastic net (alpha=0.400)	Post-est OLS
childs	0.0637454	0.0904251
agekdbrn	-0.0059233	-0.0188758
partyid	0.0060137	0.0717246
polviews	0.0104986	0.0444641
postlife	-0.0557598	0.0636586
pray	-0.0406367	-0.0749151
happy	-0.2373295	-0.6023886
kidssol	-0.0442531	-0.1089418
fechld	0.0330535	0.2534808
fefam	-0.096058	-0.031779
promteok	-0.0996091	-0.1444552
hlthdays	0.0082154	0.0281105
workfor1	0.0438553	0.203827
partners	-0.0019219	-0.2721374
Partialled-out*		
_cons	3.0339375	3.1150064

4.3 Elastic Net with $\alpha = 0.6$

Changing the code above from $alpha(0.4)$ to $alpha(0.6)$ yields the results as:

Selected	Elastic net (alpha=0.600)	Post-est OLS
childs	0.0493219	0.1610319
agekdbrn	-0.0021383	-0.0270679
postlife	-0.0311691	-0.2154722
pray	-0.0128982	-0.1064644
happy	-0.1442972	-0.4624108
fefam	-0.0547979	-0.1307711
promteok	-0.0857415	-0.1823602

Partialled-out*		
_cons	2.679081	4.5054702

4.4 Elastic Net with $\alpha = 0.8$

Changing the code above from $alpha(0.6)$ to $alpha(0.8)$ yields the results as:

Selected	Elastic net (alpha=0.800)	Post-est OLS
childs	0.0087408	0.2040651
happy	-0.0564257	-0.4788543
postlife	-0.0208348	-0.1850568
fefam	-0.0532010	-0.2316977
Partialled-out*		
_cons	2.3085033	3.5819987

5 Comparasion and Conclusion

The table below is the summary table of all these regulization models:

Selected	Elastic net	Elastic net	Elastic net	Elastic net	Lasso
	(alpha=0.200)	(alpha=0.400)	(alpha=0.600)	(alpha=0.800)	(alpha=1.000)
papres10	0.0010458				
childs	0.0680969	0.0637454	0.0493219	0.0087408	
agekdbrn	-0.0058746	-0.0059233	-0.0021383		
partyid	0.0210815	0.0060137			
polviews	0.0484773	0.0104986			
postlife	-0.0636745	-0.0557598	-0.0311691		
pray	-0.056648	-0.0406367	-0.0128982		
happy	-0.3281692	-0.2373295	-0.1442972	-0.0564257	-0.0022546
conmedic	0.0656661				

Selected	Elastic net	Elastic net	Elastic net	Elastic net	Lasso
	(alpha=0.200)	(alpha=0.400)	(alpha=0.600)	(alpha=0.800)	(alpha=1.000)
kidssol	-0.1056309	-0.0442531			
spanking	-0.0241453				
fechld	0.0757882	0.0330535			
fefam	-0.1073232	-0.096058	-0.0547979	-0.0208348	
overwork	-0.0147769				
promteok	-0.1098092	-0.0996091	-0.0857415	-0.053201	-0.0288106
hlthdays	0.016018	0.0082154			
workfor1	0.1040439	0.0438553			
partners	-0.0512729	-0.0019219			
eth1	0.0023778				
Partialled-out*					
_cons	2.9355073	3.0339375	2.679081	2.3085033	2.108893

As we can see from the above result table that both the number and the corresponding coefficient of variables being left after the regularization are getting smaller as the hyperparameter α of the elastic net getting closer and closer to 1 (finally the model changes from Elastic Net to Lasso).

References

[1] Conduct cross-validation with cvlasso *https* : *//statalasso.github.io/docs/cvlasso/*

[2] The help file of cvlasso in stata. *https* : *//statalasso.github.io/docs/cvlassohelp/*

[3] The help file of lasso2 in stata. *https* : *//statalasso.github.io/docs/lasso2help/*

[4] Jiaming MAO. "Model Selection and Regularization"

https : *//github.com/jiamingmao/data − analysis/blob/master/Lectures*

[5] GSS dataset. *http* : *//gss.norc.org*