Which factors will affect Chinese Individual's Expected Income?

Yang Yuzhou

We use CGSS data in 2006 to evaluate the effects that some factors affect Chinese Individual's expected income. First, we should deal with the original data from CGSS2006. The data contains basic demographic information and another variables about income directly. The numer of the variables is 71. Therefore, we get the final 33 variables for estimation. By the way, we put the detail code to deal with the original data in a new do file becasue markdown for Stata 14.0 cannot support Chinese and the code is too long to present all. Therefore, we have

```
. use cgss2006_14.dta, clear
.
. qui do data.do
```

Second, we use `-cvlasso-` to program for cross-validation using elastic net estimators. The key point we choose elastic net estimators is that the dependent variable whether next three years' income is expected to increased is indicators variable:

```
. global x qs3 qa01 qa02 qa04 qa05a qa08a qa10 qa11 qa13 qb01d quc321 ///
> quc322 quc323 quc324 quc325 quc326 qc34a qc34b qc492 qd06e qd07h1 ///
> qd08h qd17c qd35a qd36a qd36b qd36c qe02 qe101 qd13i qe03 qe04
.
. gen y = qe111 //Expectation Income next three years.
. sum y, de
```

y

Percentiles		Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	3,461
25%	0	0	Sum of Wgt.	3,461
50%	0		Mean	.4154869
		Largest	Std. Dev.	.492877
75%	1	1		
90%	1	1	Variance	.2429277
95%	1	1	Skewness	.3429876
99%	1	1	Kurtosis	1.117641

If $y = 1$, the individual expected the future's income would increase; if $y = 0$,

not. Thus, we should use `-cvlasso-` to find a suitable α and λ. In order to find the suitable α, we use $\alpha = 0, 0.5, 1$ to minimize the λ. For the rubustness, we use the LSE and LOPT method to estimate[1]:

```
. /*
> qui cvlasso y $x, alphacount(3) lse
> dis e(alphamin)
> dis e(lmin)
> qui cvlasso y $x, alphacount(3) lopt
> dis e(alphamin)
> dis e(lmin)
> */
```

Then we should use $\alpha = 1$ and $\lambda = 0.16124761$. It is easy to see from the graph

```
. qui cvlasso y $x, alpha(1) plotcv
. graph export CVLOPT.png, replace
(file CVLOPT.png written in PNG format)
. ![](CVLOPT.png)
```

Third, we use the optimal α and λ to program `-lasso2-` to get the result.

```
. lasso2 y $x, alpha(1) l(`e(lmin)´)
```

Selected	Lasso	Post-est OLS
qa01	0.0132576	0.0102691
qa02	-0.0067497	-0.0066664
qa04	0.0050534	0.0027365
qa05a	0.0042316	0.0032621
qa08a	0.0158816	0.0161012
qa10	0.1133575	0.1163855
qa11	-0.2761343	-0.2722991
qa13	0.0283148	0.0298432
qb01d	-0.0002592	-0.0001993
quc321	-0.0068355	-0.0097835
quc322	-0.0094091	-0.0106919
quc323	-0.0035897	-0.0031302
quc324	0.0532467	0.0508629
quc325	-0.0115487	-0.0179406
quc326	0.0216805	0.0238974
qc492	-0.0337470	-0.0370073
qd07h1	-0.0289508	-0.0314209
qe02	0.0573976	0.0523987
qe101	0.4371835	0.4368777
Partialled-out*		
_cons	0.3904537	0.3957075

Finally, we obtain the selected variables. It is a intuitively story : A woman (qa01) not young(qa02), belongs to minority(qa04), less educated(qa05a), do not take part in the community(qa08a qa10). She is a soldier (qa11), has a belief(qa13) and works longer every week(qb01d). Her total income is from the

[1] We have simulated α in smaller gap, e.g., $\alpha = 0, 0.1, 0.2, \ldots, 1$, and the results will not change.

army(quc321) which is a stable job(quc322). And the income is mainly sourced by the salary and bonus(quc323), and seldom contains dividends(quc324). If she get dividends by chance, the additional income is mainly from her higher position(quc325) and hardly from her performance(quc326). And her income had a large change from the Reform(qc492). Her husband has less income than her(qd07h1). And she doesn't satisfy the income level(qe02) which has litte change compared to the one before three years(qe101). At this moment, She would be the most impossible to expect the income to be increased.