

Exercise 3
Advanced Maximum Likelihood
ICPSR Summer Program
Charles Franklin, Instructor

The file ~~GSSTOBIT.DTA~~ contains data from the General Social Survey. A number of years have been deleted due to completely missing questions or half sampling. The remaining data range from 1975 through 1989, with some holes. Check the marginals for year to see what is there.

There are two variables of interest of political scientists, and perhaps others. One is liberal-conservative self-identification. ~~The other is confidence in the U.S. Supreme Court.~~ Both variables have some missing data, suggesting that sample selection may affect analysis of these variables. A Heckman/Type 2 Tobit model may offer insight into the selection process and its effects.

Construct a model to address the substantive question of how political views ~~and confidence in the Supreme Court~~ have changed over time, while controlling for relevant individual factors. Specify as part of this, a reasonable selection model. Use the heckman command in Stata to estimate ~~the type 2 Tobit model for these two items.~~ While neither dependent variable is in fact continuous, we will ignore that fact for the sake of the exercise.

As you construct your model, remember that missing data on the independent variables will also cause cases to be dropped from the analysis. Some of the variables you may wish to use are only available in some years. So care must be used in selecting the model in light of the constraints imposed by the data. This can be frustrating, but it is also reality.

The variables are reasonably self-documenting with value labels. Here is the list of variables and their descriptive statistics:

```
. desc
```

```
Contains data from gsstobit.dta
```

```
obs:      14,152
vars:      23                23 Jul 1999 09:48
size:      424,560 (99.3% of memory free)
```

```
-----
 1. year      byte   %4.0g      gss year for this respondent
 2. id        int    %8.0g      respondent id number
 3. marital   byte   %4.0g      marital status
 4. age       byte   %4.0g      age of respondent
 5. educ      byte   %4.0g      highest year of school complete
 6. degree    byte   %4.0g      rs highest degree
 7. sex       byte   %4.0g      respondents sex
 8. race      byte   %4.0g      race of respondent
 9. income    byte   %4.0g      total family income
```

```

10. region    byte    %4.0g    region of interview
11. size      int     %8.0g    size of place in 1000s
12. partyid   byte    %4.0g    political party affiliation
13. polviews  byte    %4.0g    think of self as liberal or con
14. relig     byte    %4.0g    rs religious preference
15. union     byte    %4.0g    does r or spouse belong to unio
16. news      byte    %4.0g    how often does r read newspaper
17. talkpol1  byte    %4.0g    r discusses politics with perso
18. talkpol2  byte    %4.0g    r discusses politics with perso
19. talkpol3  byte    %4.0g    r discusses politics with perso
20. conjmiss  byte    %9.0g    Conjjudge missing=0,observed=1
21. polvmiss  byte    %9.0g    Polviews missing=0,observed=1
22. yearsq    int     %9.0g    Year squared
23. union2    byte    %9.0g    Union house=1, not=0

```

Sorted by:

```
. summ
```

Variable	Obs	Mean	Std. Dev.	Min	Max
year	14152	79.6657	4.467752	74	87
id	14152	862.4338	595.2974	1	3247
marital	14150	2.039788	1.518296	1	5
age	14078	44.78463	17.75445	18	89
educ	14100	11.93851	3.218123	0	20
degree	14111	1.055063	1.082435	0	4
sex	14152	1.562959	.4960378	1	2
race	14152	1.178985	.4148701	1	3
income	13518	8.79568	3.18133	1	13
region	14152	4.681388	2.400009	1	9
size	14152	414.4201	1339.872	0	7895
partyid	13942	2.449577	1.944353	0	6
polviews	13344	4.050659	1.318094	1	7
relig	14111	1.548508	.93641	1	5
union	9227	3.439363	1.089672	1	4
news	9676	1.844771	1.194598	1	5
talkpol1	1689	3.342806	1.651603	1	6
talkpol2	1405	3.669039	1.546318	1	6
talkpol3	1052	3.715779	1.533023	1	6
conjmiss	14152	.9515263	.2147726	0	1
polvmiss	14152	.9429056	.2320311	0	1
yearsq	14152	6366.583	721.256	5476	7569
union2	9227	.2230411	.4163082	0	1