

Panel GLM Lab

May 11, 2015

We will be working with a somewhat famous political behavior dataset called “The Political Socialization Panel,” which is a multi-generational, multi-year survey research project begun in 1965 by M. Kent Jennings and Richard G. Niemi. They originally interviewed graduating high school students and their parents in 1965. They then re-interviewed the students in 1973, 1982, and 1997. Numerous books and articles have been published on the data by these authors and their collaborators on topics ranging from political participation, tolerance, opinion change, parent-child political socialization, partisanship, ideology, and so forth.

The data are useful to us this week because they provide a four-wave panel dataset that includes numerous binary, ordered, and count variables to analyze. You can obtain the data from Blackboard at **Additional materials > Data > SocializationPanel.dta**. These data are in *wide* format, so that folder also contains a do-file (**SocializationPanel.do**) that relabels the variables and restructures the data to *long*. A codebook for the original, wide dataset is also provided (**SocializationPanel.pdf**).

1 Getting Started

1. Before working with these data, you may want to work through Cameron and Trivedi Ch.18 to get a sense of what’s about to happen.
2. Load the data file into Stata.
3. Open and run **SocializationPanel.do**. You can add the following analyses to that file.
4. Use **xtset** to specify **respondent** and **year** as the features of the panel structure. Recall you can use **tab** (by **year**), **xttab**, **xttrans**, **xtdescribe**, and **xtsum** to obtain various summaries of panel data.

2 Binary Outcomes

5. We will start by trying to model whether a respondent has a university degree (see `edudegree`). This variable is categorical and missing values mean no *new* university degree since the last panel interview wave. Create a new variable `degree` that is 1 for anyone who has a degree from the current or any previous wave and 0 in all waves where they have not yet earned a degree. For someone who never earns a degree, all of their values 0; for someone who earns a degree before 1973, all of their values are 1; for someone who earns a degree between 1973 and 1982, their value is 0 in 1973 and 1 thereafter. Note that no one has a degree in 1965. You will have to reshape the data back to wide to do this, create year-specific new variables (named following the 65, 73, 82, 97 convention) and then use `reshape long` to return to long format, but you will have to add your new variable to the varlist in that command:

```
reshape wide
gen degree65 = 0
recode edudegree73 (.=0) (else=1), gen(degree73)
recode edudegree82 (.=0) (else=1), gen(degree82)
recode edudegree97 (.=0) (else=1), gen(degree97)
replace degree82 = 1 if degree73 == 1
replace degree97 = 1 if degree82 == 1
reshape long educ edudegree partyid spedu spid ideology intefficacy ///
             extefficacy ptrust trust trustgov knowledge tola tolb ///
             working indexa indexb polinfo sex race class relig ///
             home incomef incomep thunion thbus thwhite thblack ///
             state degree, i(respondent) j(year)
```

6. Now look at `tab degree year` to see the development of education.¹
7. Use `logit` and `probit` to obtain pooled estimates (ignoring the panel structure) of the effect of family income (`incomef`, recoded to drop 0 and 97 values) and any other variables of your choice on the outcome. Use clustered standard errors.
8. Interpret the effects using `margins`. You can also obtain pooled estimates using `xtlogit` or `xtprobit` with the `, pa` option. Coefficients may differ to some extent from what you just estimated.
9. Now obtain fixed and random effects estimates using `xtlogit` with the `, fe` and `, re` options.
10. Attempt to interpret these results using the following approaches:
 - Use `margins`, `predict(xb)` to obtain predicted log-odds

¹Note: You could also try analyzing these data as an event history by returning the data to wide format and creating a new variable specifying the time at which each person earned their degree.

- Use `margins, dydx(*) predict(pu0)` to obtain predicted probabilities, assuming the fixed or random effect is zero
- Use `margins, dydx(*) predict(xb)` to obtain marginal effects in log-odds terms
- Use `margins, dydx(*) predict(pu0)` to obtain marginal effects in probability terms, assuming the fixed or random effect is zero

See discussion in Cameron and Trivedi (p.630) or `help xtlogit postestimation` about the options for interpretation. In general, our usual fallback on `margins` does not provide the same straightforward interpretation as in cross-sectional analyses.

11. Use a Hausman test to decide whether to use fixed or random effects (i.e., estimate using random effects, then fixed effects, storing coefficients each time, and then perform `hausman` on the stored estimates). The null hypothesis being tested is that the fixed effects are uncorrelated with other covariates. Rejecting it means that we should use the fixed effects approach.
12. It is not possible to estimate a fixed-effects probit model, but you can estimate a random effects model using `xtprobit`, `re`. Try this and attempt to interpret the results.
13. Recall you can also estimate an event-history model for the `degree` variable because it only changes once for each respondent. Feel free to apply methods from last week to achieve this, if you so desire.
14. You can now try any of these procedures again for other binary outcome variables, such as whether someone owns their own home (`home`) or whether a respondent thinks most other people can be trust (`trust`).

3 Ordered Outcomes

15. There are several variables in the dataset that might be considered ordinal, such as `tola` and `tolb` (items about one's level of political tolerance), `trustgov` (a three-item measure of trust in the government), etc. You can use whichever variables you think can be treated ordinally to perform these analyses. We'll focus on the `trustgov`.
16. `trustgov` has three categories coded 1,3,5. Create a new measure, `trustgov2`, that ranges from 0 to 2:

```
recode trustgov (1=2) (3=1) (5=0), gen(trustgovb)
```
17. How do levels of trust in the same vary over time?
18. Performed pooled OLS and pooled ordered logit and probit analyses on these data, with whatever putatively causal variables you so choose. Interpret using `margins`.

Specify a dummy variable for each respondent to manually obtain fixed effects estimates (use respondent-clustered standard errors in this case).

19. Now, respect the panel data structure and use `xtreg` to obtain estimates and interpret the results in substantive terms using `margins`.
20. We can obtain random effects estimates using `xtlogit` and `xtoprobit`. These commands only estimate random effects models, so you don't need to specify `, re`.
21. Attempt to make substantive interpretations using the `margins` command with the options described for binary outcome models.²
22. Repeat this process looking at the measures of tolerance. Assess whether the difference between the measures is substantively consequential.

4 Count Outcomes

23. There are three variables in the dataset that are explicitly counts: `knowledge`, `indexa`, and `indexb`. The latter two are measures of numbers of participatory events (i.e., how many acts of political participation the respondent has engaged in at each panel wave).
24. `knowledge` is coded 1–7, where 1 means zero correct answers to questions about political information and 7 represents 6 correct answers. Use `replace` to recode `knowledge` to range from 0–6.
25. Begin by estimating a simple pooled OLS model, with `knowledge` as a function of gender (you should create a new variable `female` that is 1 for women and 0 for men). How much more knowledgeable do men appear to be than women?
26. Now estimate a fixed-effects model using `xtreg` with the `, fe` option. Why can't you estimate the effect of gender?
27. Now estimate the model using a random effect specification. What is the effect of gender on `knowledge`?
28. Now model this in a Poisson framework (i.e., use `poisson` to obtain a pooled estimate that ignores panel structure). Keep in mind the need for clustered standard errors. Use `margins` to interpret the effect of gender. How much less knowledgeable are men than women? How does this effect vary across years, if at all?
29. A concern with Poisson is overdispersion. Test for this (e.g., using the test described on p.575 of Cameron and Trivedi). Does there seem to be overdispersion? If so, what do you do?

²You will need to move to a linear outcome model, mixed effects estimation, or a pooled model in order to make other kind of substantive interpretations.

30. One strategy for dealing with overdispersion is to use negative-binomial regression. Again, estimate a pooled model ignoring panel structure using the `nbreg` command. How much less knowledgeable are women than men? Is that something that can be readily interpreted from the coefficient estimates?
31. Now, re-estimate these models (OLS, Poisson, and negative-binomial) respecting the panel structure.
 - Start with OLS using a random effects specification: `xtreg knowledge female, vce(robust)`. Interpret the effect of gender using `margins`.
 - Re-estimate using Poisson: `xtpoisson knowledge female, vce(robust)`. Interpret using `margins`. What's going on here? The default behavior of `margins` maybe unexpected; try again using the `, predict(nu0)` option.
 - Re-estimate using negative-binomial (with `xtnbreg`): `xtnbreg knowledge female`. What happens? Why?
32. Okay, okay, the above results have been hard on women. Let's focus on something else and, in particular, let's focus on a variable that is time-varying at the unit-level. Estimate a model that looks for the effect of education on knowledge. The education variable `educ` is categorical; recode it so that it represents any education (all categories are 1 and missing is recoded to 0).
33. Use a pooled specification, fixed effect, and random effects and use OLS, Poisson, and negative-binomial alternatives. Compare the results.

5 Multilevel Approach

34. Recall that panel data are a particular kind of hierarchical data structure where units are observed within time periods and the same cases are observed multiple times. (This differs from a repeated cross-sectional data structure, which is also hierarchical with units observed within time periods but where units are only observed once.) You could use a mixed effects, hierarchical regression approach to analyze panel data.
35. If you're interested, you can examine the functions `melogit` and `meprobit` (for binary outcomes), `meologit` and `meoprobit` (for ordered outcomes), and `mepoisson` and `menbreg` (for count outcomes), and `mixed` (for continuous outcomes). You're welcome to pursue any of these approaches on your own. The Stata Multilevel Mixed-Effects Reference Manual (<http://www.stata.com/manuals13/me.pdf>) provides some guidance.