Unit 3
Longitudinal and Multilevel Models
3-2: Models for Ordered, Nominal and Count Outcomes

PS2730-2021

Week 11

Professor Steven Finkel

Including Unit-Level Heterogeneity in Ordered, Nominal and Count Outcomes

- Choices for incorporating unit-level heterogeneity are generally the same as those we covered for dichotomous dependent variables: random effects, fixed effects, and the random effects hybrid model
- As always, the choice between estimating a pure random effects model versus the other twodepends on whether the assumption of no X_{it}-ζ_i correlation is tenable. If so, we prefer RE models as they are most efficient.
- If not (or if a direct comparison between the models shows violation of the assumption), then the choice is between fixed effects and the hybrid model
- If a fixed effects model is possible to be estimated (which is not always the case), it is done either via a variant of the Chamberlain conditional maximum likelihood method, or (if permitted) by including N-1 dummy variables without the problem of incidental parameters alluded to earlier
- In virtually all instances, the hybrid model works well

Ordered Logit and Ordered Probit

- How can we incorporate unit-level heterogeneity in models for ordinal outcomes?
- There is no fixed effects ordered logit -- you can't condition out the ζ like in the dichotomous case by summing the dependent variable, nor can you add dummy variables, given the incidental parameters problem
- The best solution is the hybrid model, which can easily be implemented in Stata with the options we considered for dichotomous outcomes:
  - xtologit (two-level basic ordinal regression with random unit-level intercept)
  - meologit/meoprobit (two or multilevel ordinal regression with random unit-level intercept and/or random coefficients at multiple levels)
  - GSEM
- There are also ways (and a Stata module "Extended Regression Models") to incorporate other forms of endogeneity into these models. For example, we can model sample selection, or time-varying correlation of unobservables with the X covariates. We'll cover this in the next unit.

- Ordered Logit:

$$P(Y_{it} > m \mid X_{it}, \overline{X}_i, \varsigma_i) = \frac{\exp^{X_{it}\mathrm{B} + \overline{X}_i\mathrm{B} + \varsigma_i - \tau_m}}{1 + \exp^{X_{it}\mathrm{B} + \overline{X}_i\mathrm{B} + \varsigma_i - \tau_m}}$$

- Where the **τ** are the cut-points for the ordered categories, ζ is the unit-level heterogeneity parameter, X_{it} are the time-varying variables, and the X-bar are the unit-level time-invariant means for the time-varying variables
- This model is a non-linear cumulative probability model predicting being *above* category m (as opposed to our earlier treatment of the cumulative probability of being at or *below* category m – this is just an algebraic convenience)

- So cumulative odds of being above versus being at or below category m are:

$$\frac{P(Y > m \mid x, \varsigma)}{P(Y \le m \mid x, \varsigma)} = \exp^{X_{it}\mathrm{B} + \overline{X}_i\mathrm{B} + \varsigma_i - \tau_m}$$

- And cumulative log-odds linearizes the model as:

$$\ln \frac{P(Y > m \mid x, \varsigma)}{P(Y \le m \mid x, \varsigma)} = X_{it}\mathrm{B} + \overline{X}_i\mathrm{B} + \varsigma_i - \tau_m$$

- Estimated via ML methods similar to the complexities encountered in xtlogit given the additional normally distributed ζ term
- As in ologit, the model assumes "proportional odds" in that the Betas are the same for all cumulative probability categories
- The Betas for the X_{it} are the time-specific effects of the X on the cumulative log-odds, the Betas for the X-bars are the "between effects" representing the propensity of units to be "higher" or "lower" on the outcome based on their average level of X (and all stable unobservables correlated with X)
- Test of the equality of the "within" and "between" effects (when expressed in mean-deviated form) tests whether assumption of naive RE is appropriate

- We can derive probabilities of being in each category

$P(Y=1|X_{it},\zeta_i) = 1 - \frac{expX_{it}B+\bar{X}_iB+\zeta_i-\tau_1}{1+expX_{it}B+\bar{X}_iB+\zeta_i-\tau_1}$ or 1-cumulative P(Y>1)

$$P(Y=2|X_{it},\zeta_i) = \frac{expX_{it}B+\bar{X}_iB+\zeta_i-\tau_1}{1+expX_{it}B+\bar{X}_iB+\zeta_i-\tau_1} - \frac{expX_{it}B+\bar{X}_iB+\zeta_i-\tau_2}{1+expX_{it}B+\bar{X}_iB+\zeta_i-\tau_2}$$

or cumulative P(Y>1) minus cumulative P(Y>2)

$$P(Y=3|X_{it},\zeta_i) = \frac{expX_{it}B + \bar{X}_iB + \zeta_i - \tau_2}{1 + expX_{it}B + \bar{X}_iB + \zeta_i - \tau_2}$$

or cumulative(P>2)

- Estimate β and ψ (variance of the ζ) via ML methods which maximize the joint probability of observing the 1, 2, 3s we observed in the sample
- Be certain to estimate standard errors clustered by unit

- Ordered Probit
- Same procedure for extending the ordered probit model to arrive at the multilevel/longitudinal version

$$Y_{it}^{*} = \Sigma\beta X_{it} + \Sigma\beta \overline{X}_{i} + \varsigma_{i} + \varepsilon_{i}$$

$$Y_{it}^{*} = X\mathrm{B} + \overline{X}\mathrm{B} + \varsigma_{i} + \varepsilon_{i}$$

$$\mathrm{E}(Y_{it}^{*} \mid X) = X\mathrm{B}$$

- Y* is continuous but unobserved. We map the observed variable Y_i to Y* via the "measurement equation" that says if Y* is above a certain threshold, observed Y will be 1; if Y* is above the next threshold, observed Y will be 2; above the next threshold, observed Y will be 3, and so on, depending on the number of categories

- Assume a three category ordinal variable
- Assign the zero threshold (τ_0) to be negative infinity ($-\infty$) and the threshold for the last category (τ_3) to be positive infinity (∞)
- Then the full model is:

$$Y_{it}^* = \Sigma\beta X_{it} + \Sigma\beta \overline{X}_i + \varsigma_i + \varepsilon_i$$
$$Y_{it}^* = X\mathrm{B} + \overline{X}\mathrm{B} + \varsigma_i + \varepsilon_i$$
$$\mathrm{E}(Y_{it}^* \mid X) = X\mathrm{B}$$

with measurement equations

$$Y_i = 1 \text{ if } \tau_0(-\propto) <= \mathrm{Y}_i^* < \tau_1$$
$$Y_i = 2 \text{ if } \tau_1 <= \mathrm{Y}_i^* < \tau_2$$
$$Y_i = 3 \text{ if } \tau_2 <= \mathrm{Y}_i^* < \tau_3(\propto)$$

- And we can work out the probabilities of being in each category

$$P(Y=1 \mid X, \overline{X}, \varsigma) = P(\tau_0 \le Y^* < \tau_1)$$
$$P(Y=1 \mid X, \overline{X}, \varsigma) = P(\tau_0 \le X\mathrm{B} + \overline{X}\mathrm{B} + \varsigma_i + \varepsilon_i < \tau_1)$$
$$P(Y=1 \mid X, \overline{X}, \varsigma) = P(\tau_0 - X\mathrm{B} - \overline{X}\mathrm{B} - \varsigma_i \le \varepsilon_i < X\mathrm{B} + \overline{X}\mathrm{B} + \varsigma_i)$$
$$P(Y=1 \mid X, \overline{X}, \varsigma) = P(\varepsilon_i < X\mathrm{B} + \overline{X}\mathrm{B} + \varsigma_i)$$

(because τ_0 is negative infinity)

$$P(Y=2 \mid X, \overline{X}, \varsigma) = P(\tau_2 - X\mathrm{B} - \overline{X}\mathrm{B} - \varsigma_i) \le \varepsilon_i < \tau_1 - X\mathrm{B} - \overline{X}\mathrm{B} - \varsigma_i)$$
$$P(Y=3 \mid X, \overline{X}, \varsigma) = P(\tau_2 - X\mathrm{B} - \overline{X}\mathrm{B} - \varsigma_i) \le \varepsilon_i$$

- And, assuming normally distributed ε, we arrive at the generalized ordered probit model as:

$$P(Y=m \mid X, \overline{X}, \varsigma) = \Phi(\tau_m - X\mathrm{B} - \overline{X}\mathrm{B} - \varsigma_i) - \Phi(\tau_{m-1} - X\mathrm{B} - \overline{X}\mathrm{B} - \varsigma_i)$$

Nominal Outcomes

- How can we incorporate unit-level heterogeneity in models for nominal outcomes?
- Typically implemented via multilevel multinomial logit – in Stata as XTMLOGIT as of version 17
- Idea is that there is a unit-level propensity to select a given outcome category *j* from the set of *J* outcomes at all times ζ_{ij}, regardless of the levels of X_{it}. So there will be J-1 heterogeneity ζ_{ij} terms which may be estimated in the context of the random, fixed, or hybrid models we've been considering

Random Utility (Choice) Models and MNL extension

- Can derive this model from the random utility (RUM) or choice models we used to motivate the MNL model in the non-longitudinal/multilevel case
- Utility for individual *i* for choice *j* at time *t* is a function of both observed factors (X_{it}), unobserved subject-choice factors (ζ_{ij}), and unobserved subject-choice-time factors (ε_{ijt}):

$$U_{ijt} = X_{it}\beta_j + \zeta_{ij} + \varepsilon_{ijt}$$

where the ζ_{ij} is what we have defined earlier – the unobserved factors that push a given individual's utility up or down for a given choice regardless of the time period

- Assuming a "Type 1 extreme value distribution" for the ε_{ijt} and Independence of Irrelevant Alternatives (IIA) yields the longitudinal multinomial logit model

- One category serves as the baseline against which the utility of a given choice is compared
- The probability of selecting category *j*:

$$P(y = j|x, \text{not baseline category } b) = \frac{exp^{XB_j + \zeta_{ij}}}{\sum exp^{XB + \zeta_{ij}}{}_{j \neq b} + 1}$$

$$P(y = j|x, \text{baseline category } b) = \frac{1}{\sum exp^{XB + \zeta_{ij}}{}_{j \neq b} + 1}$$

- The **Random Effects Multinomial Logit Model** assumes that the ζ_{ij} are normally distributed and ML estimation integrates them out in the same way as in multilevel/longitudinal RE logit and RE ordered logit
- The ζ_{ij} are assumed to be unrelated to the X_{ijt} **but** in the multinomial case they **may** be correlated with each other – so the unobserved factors that give an individual utility for choice j_1 over the baseline may be related to the same unobserved factors that give an individual utility for choice j_2 over the baseline category
- Model becomes unwieldy as the number of *J* choice categories increases

- As always, if $E(X_{ijt}\zeta_{ij}) \neq 0$, then move to alternatives – Fixed Effects or Hybrid models
- One (inferior) possibility for Fixed Effects MNL would be to estimate a series of bivariate fixed effects logit models using the same conditional effects (Chamberlain) method we discussed earlier, selecting one category as the baseline outcome.
  - Benefit: straightforward and easy to implement, no specialized software needed
  - Drawbacks: results may differ depending on which category is chosen as the baseline; no test of overall effect of a variable on the outcome; will produce results based on differing sample sizes depending on missing data in a given comparison of outcomes
- Better: extend the Chamberlain method for FE estimation within a single model covering all choices over time

- **Fixed Effects Multinomial Logit**: add the ζ_{ij} heterogeneity term to each outcome choice for a given individual, and then, condition the heterogeneity term out of consideration by maximizing the ***conditional likelihood*** for the outcomes, given the **total sum of the choices** for that category by that individual over time
- For example, in a three-wave situation, individual i chooses alternative j_1 at time 1 and j_2 at times 2 and 3. Her choice sequence is therefore {1,2,2}. The conditional effects model says: of all the possible permutations with choices 1 and 2, what was the likelihood of the sequence {1,2,2}? Note that the universe of sequences with 1 and 2 is {1,2,2}, {2,1,2}, (2,2,1}. So the "sum" of choices over time for alternative 1 is 1 and alternative 2 is 2. This procedure conditions on the sum of the choices being 1 for alternative 1, 2 for alternative 2.
- That leaves three ways of getting the sum of "1" for alternative 1 over time; what was the likelihood of choosing alternative 1 in wave 1 compared to wave 2 or 3?
- Similarly: 3 ways of getting the sum of 2 for alternative 2: waves 2 and 3, 1 and 3 and 1 and 2. What was the likelihood of choosing alternative 2 in waves 2 and 3?
- All of this conditions out the ζ_{ij} from consideration, and hence any correlation between X_{ijt} and ζ_{ij} is irrelevant to the estimation!

Drawbacks:

- Lose cases, as in dichotomous conditional fixed effects model, where the alternative chosen is the same at every point in time
- Computationally intensive
- Becomes unwieldy as the number of time points (waves) increases. More waves means more possible permutations to get a sum of "1" or "2" or whatever value for a given alternative – when T gets even moderately large the permutations explode exponentially. For example, in a panel with T=15 and 6 alternatives where a single individual chooses alternative 1 four times, there are over 375 million permutations of "4" choices of 1 over the 15 waves. Crazy!
- Solution is to take a "random sample" of the permutations in the analysis set but this solution is (obviously) not maximally efficient

Random Effects Hybrid Multinomial Logit Model

- Construct X-bar for each time-varying independent variable, and follow the general procedures we've discussed for the RE hybrid model
- Use XTMLOGIT or GSEM for estimation
- Benefits:
  - uses all cases, not only cases which changed on the outcome choice over time
  - Allows testing of the equality of the "within" and "between" effects via comparison of the regression coefficients for X_{it} and X-bar_i (in the mean-deviated form of the model)
  - Allows all other advantages of the multilevel/longitudinal set-up discussed earlier
- Be certain to estimate robust standard errors, clustered by unit

Extension: Count Outcomes

- How can we incorporate unit-level heterogeneity in models for count dependent variables?
- Recall the Poisson model for non-negative integer counts:

$$\Pr(y \mid \mu) = \frac{\exp^{-\mu}(\mu^{y})}{y!}$$

- With rate parameter μ being an exponentiated function of the independent variables X

$$\mu = E(y \mid X) = \exp(XB)$$

- With longitudinal data we have time-varying Xs which then lead to different rates at different times
- But we can also incorporate a unit-level heterogeneity term ζ_i, which pushes counts up or down for that unit as well at all points in time

$$\mu_{it} = E(y_{it} \mid X_{it}, \varsigma_i) = \exp(X_{it}B + \varsigma_i)$$

- This then translates into probabilities of particular counts at particular times according to the Poisson distribution

$$\Pr(y_{it} \mid X_{it}, \varsigma_i) = \frac{\exp(-\mu_{it})(\mu_{it}{}^{y})}{y!} = \frac{\exp(-\exp(X_{it}B + \varsigma_i))(\exp(X_{it}B + \varsigma_i))^{y}}{y!}$$

where $\mu_{it} = \exp(X_{it}B + \varsigma_i)$

- The model with "fixed effects" for the ζ_I is estimated via another variant of the "conditional maximum likelihood" Chamberlain method used in FE logit and FE Multinomial Logit
- The probability of counts at a particular time, *given the total sum of counts for the unit over time*, can be estimated via ML methods without consideration of the ζ_I , and hence effects are estimated regardless of the correlation of the Xs and the ζ_I – exactly what we want FE to do!
- This is estimated in Stata via "xtpoisson" with fe option

- However, xtpoisson typically suffers from the same problem as regular poisson regression, which is the issue of "overdispersion" where the variance of the counts does not equal the mean
- We handled that problem with a negative binomial regression which added variance via a term to be estimated (which we called "alpha"); this term was assumed to follow a gamma distribution
- The alpha term was *also* described as a heterogeneity term, such that individuals with the same Xs at a given point in time were estimated to have different rates. You can think of this kind of heterogeneity as "time-specific" while the heterogeneity represented by ζ_i is "unit-specific" or "time-invariant".
- This is analogous to the effects of stable unobservables (ζ_i) versus unstable (time-specific) unobservables (ε_{it}) in general longitudinal models
- Technically, in the NB longitudinal case, we estimate an alpha parameter which allows for time-specific heterogeneity for all units at a given XB, and ζ_i is then added on top of this to generate the expected counts for the individual as well as the variances in the counts across individuals

- Can estimate the generalized negative binomial model in several ways:
- FE version: add dummy variables for all cases to estimate ζ_i and estimate a regular "nbreg" model with standard errors clustered by unit
- Drawback: potential incidental parameters problem mentioned earlier, but Allison (p.64) shows that the incidental parameters issue does not lead to systematic bias in this case
- Additional drawback: as with all non-linear FE models, only units that change at some point on the DV are included in the analyses; units that are always on the same count contribute nothing to the likelihood function and are discarded from consideration

- Alternative: hybrid random effects negative binomial model which includes X_{it} as well as the X-bar$_i$ as explanatory variables

- As with all RE hybrid models, we obtain the effects of time-specific values of X on the unit's time-specific rate μ_{it} , what we can call the "within unit" effect of X, as well as the effect of the unit's average on X on the average rate for the unit at all points in time, what we can call the "between unit" effect of X
- Advantages:
  - as in all RE hybrid models, we use all cases, not only cases that change on the counts over time
  - Allows testing of the equality of the "within" and "between" effects via comparison of the regression coefficients for X_{it} and X-bar$_i$
  - Allows all other advantages of the multilevel/longitudinal set-up discussed earlier