Advanced Topics in Statistical Methods

PLSC 504

Models for Binary and Event Count Responses in Panel Data
October 31 & November 2, 2017

Introduction

Today we'll talk about models for binary and other "ugly" response variables in the panel/TSCS context. We'll discuss fixed-effects models, random-effects models, and the connection between these models and duration models. Next week we'll talk about GEEs.

A Quick Refresher on Binary-Response Logit and Probit

The latent variable approach treats dichotomous dependent variables as essentially a problem of measurement. That is, it starts with the idea that there is a continuous underlying variable in which we are interested, but we are unable to measure it. Instead, we have a dichotomous indicator of that underlying (latent) variable. Call the latent variable Y^*; the underlying model is:

$$Y_i^* = \mathbf{X}_i\beta + u_i \tag{1}$$

This model has the usual OLS-type assumptions; in particular, that u_i is randomly-distributed according to some known distribution (e.g., Normal). However, we observe only the following realization of Y^*:

$$\begin{aligned} Y_i &= 0 \text{ if } Y_i^* \leq 0 \\ Y_i &= 1 \text{ if } Y_i^* > 0 \end{aligned}$$

So, we can write:

$$\begin{aligned} \Pr(Y_i = 1) &= \Pr(Y_i^* > 0) \\ &= \Pr(\mathbf{X}_i\beta + u_i > 0) \\ &= \Pr(u_i > -\mathbf{X}_i\beta) \\ &= \Pr(u_i \leq \mathbf{X}_i\beta) \end{aligned}$$

where the last equality holds because of the symmetry of the distribution of the u_is. In other words, $Y = 1$ if the "random part" is less than (or equal to) the "systematic part." How do we figure out this probability?

If we assume that u follows some distribution, we could integrate over that distribution to get an idea of the probability that u_i fell above some point (e.g. $\mathbf{X}_i\beta$). And, in fact, this is exactly what we do. In particular, if we assume that the u_is follow a standard logistic distribution, we get a logit model:

$$\Pr(u) \equiv \lambda(u) = \frac{\exp(u)}{[1 + \exp(u)]^2} \tag{2}$$

That's the PDF, the term for the probability that u takes on some specific value. If we want to know the probability that (e.g.) u is less than or equal to some value, we consider the cumulative distribution function of the logit, which is:

$$\Lambda(u) = \int \lambda(u)du = \frac{\exp(u)}{1 + \exp(u)} \tag{3}$$

This function represents the probability that a variable distributed as standard logistic will be above some value u. It defines the familiar S-shape of a logit curve. Once we've made this assumption, we can write:

$$
\begin{aligned}
\Pr(Y=1) &= \Pr(Y^* > 0) \\
&= \Pr(u_i \leq \mathbf{X}_i\beta) \\
&= \frac{\exp(\mathbf{X}_i\beta)}{1+\exp(\mathbf{X}_i\beta)} \qquad (4)
\end{aligned}
$$

This is the basic form of the probability for the logit model. To get a probability statement for every observation i in our data, we want to think of the probability of getting a zero (one) given the values of the covariates and the parameters. That is, the likelihood for a given observation i is:

$$
L_i = \left(\frac{\exp(\mathbf{X}_i\beta)}{1+\exp(\mathbf{X}_i\beta)}\right)^{Y_i}\left[1-\left(\frac{\exp(\mathbf{X}_i\beta)}{1+\exp(\mathbf{X}_i\beta)}\right)\right]^{1-Y_i} \qquad (5)
$$

That is, observations with $Y = 1$ contribute $\Pr(Y_i = 1|\mathbf{X}_i)$ to the likelihood, while those for which $Y = 0$ contribute $\Pr(Y_i = 0|\mathbf{X}_i)$. Assuming independent observations, we can take the product over the N observations in our data to get the overall likelihood:

$$
L = \prod_{i=1}^{N}\left(\frac{\exp(\mathbf{X}_i\beta)}{1+\exp(\mathbf{X}_i\beta)}\right)^{Y_i}\left[1-\left(\frac{\exp(\mathbf{X}_i\beta)}{1+\exp(\mathbf{X}_i\beta)}\right)\right]^{1-Y_i} \qquad (6)
$$

Taking the natural logarithm of this yields:

$$
\ln L = \sum_{i=1}^{N} Y_i \ln\left(\frac{\exp(\mathbf{X}_i\beta)}{1+\exp(\mathbf{X}_i\beta)}\right) + (1-Y_i)\ln\left[1-\left(\frac{\exp(\mathbf{X}_i\beta)}{1+\exp\mathbf{X}_i\beta)}\right)\right] \qquad (7)
$$

We can then maximize this log-likelihood with respect to the $\hat{\beta}$s to obtain our MLEs. Alternatively, if we assume that our u_is follow a *standard normal* distribution:

$$
\phi(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right) \qquad (8)
$$

then the cumulative distribution function (*cdf*) is:

$$
\Phi(u) = \int_{-\infty}^{u}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du \qquad (9)
$$

which yields the *probit* model:

$$
\begin{aligned}
\Pr(Y_i=1) &= \Phi(\mathbf{X}_i\beta) \qquad (10) \\
&= \int_{-\infty}^{\mathbf{X}_i\beta}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(\mathbf{X}_i\beta)^2}{2}\right)d\mathbf{X}_i\beta.
\end{aligned}
$$

The corresponding (log)likelihood is:

$$
\ln L = \sum_{i=1}^{N} Y_i \ln\Phi(\mathbf{X}_i\beta) + (1-Y_i)\ln\Phi(\mathbf{X}_i\beta) \qquad (11)
$$

Note a few things about both of these two models:

- Each has three necessary identifying assumptions:
  1. That the "threshold" point for going from $Y = 0$ to $Y = 1$ is $Y^* > 0$. In this context, this is not a big deal, provided we include an intercept in the model.
  2. That the conditional mean of the errors is zero; i.e. $\mathrm{E}(u_i|\mathbf{X}, \beta) = 0$ (again, this is not a problem provided one has an intercept in the model).
  3. That the variances are either $\frac{\pi^2}{3}$ or one (for logit and probit, respectively).

  Of these, only the last assumption is ever particularly problematic
- Either can be motivated through a random utility model of individual choice, of the form:

$$\begin{aligned} Y_i &= A \text{ if } \mathrm{E}[U(A)] > \mathrm{E}[U(SQ)] \\ &= SQ \text{ if } \mathrm{E}[U(SQ)] > \mathrm{E}[U(A)] \end{aligned}$$

  with

$$\begin{aligned} \mathrm{E}[U(A)] &= \mathbf{X}_{iA}\beta + u_{iA} \\ \mathrm{E}[U(SQ)] &= 0 \end{aligned}$$

  so that

$$\begin{aligned} \Pr(Y = A) &= \Pr\{\mathrm{E}[U(A)] > \mathrm{E}[U(SQ)]\} \\ &= \Pr\{(\mathbf{X}_{iA}\beta + u_{iA}) > 0\}. \end{aligned} \tag{12}$$

Finally, there is the *complementary log-log* ("c-log-log") model:

$$\Pr(Y_i = 1) = 1 - \exp[-\exp(\mathbf{X}_i\beta)], \tag{13}$$

which has a likelihood and log-likelihood similar to those in (7) and (11), above. An important difference, however, is that – unlike logit and probit – the c-log-log response curve is asymmetrical, rising slowly at first, and then increasing more steeply in X. The three CDFs for these models are presented in Figure .

Panel/TSCS Data Issues

Its no big surprise that panel/TSCS models for binary dependent variables introduce some complications vis-á-vis the continuous-variable models we've discussed so far. We'll start with the basic binary-dependent-variable setup, as a latent-variable formulation:

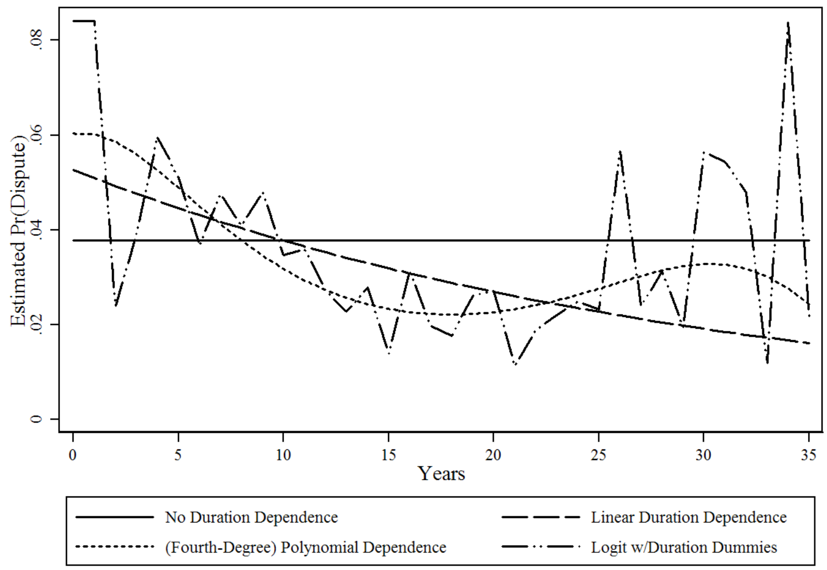
$$\begin{aligned} Y^*_{it} &= \mathbf{X}_{it}\beta + u_{it} \\ Y_{it} &= 0 \text{ if } Y^*_{it} \leq 0 \,; \\ &= 1 \text{ if } Y^*_{it} > 0 \end{aligned} \tag{14}$$

so that I'll usually write

$$Y_{it} = f(\mathbf{X}_{it}\beta + u_{it}) \tag{15}$$

where the $f(\cdot)$ indicates either a logit or probit functional form (which one really doesn't matter, at least not yet ... I'll let you know when it does). This general models is no different than a standard logit/probit setup, except that we now have $\mathbf{X}$ and Y indexed by both i and t.

Figure 1: Probit, Logit, and C-Log-Log CDFs



What Can Go Wrong?

The potential problems with panel/TSCS data in the binary context are similar to those for continuous variables: if there is unaccounted-for dependence across observations, it can mess up your estimates. Poirier and Ruud (1988) show that, when probit errors are correlated, ordinary probit still produces consistent estimates of the βs (albeit inefficient ones, and with incorrect standard errors); this is similar to the same result for standard time-series with autocorrelated errors. Beck and Katz (1997) simulate data that fits (14) with:

$$\begin{aligned} X_{it} &= \rho_X \mathbf{X}_{it-1} + \nu_{it} \\ u_{it} &= \rho_u u_{it-1} + \epsilon_{it} \end{aligned}$$

and found that, when the ρs were high, plain-vanilla logit and probit models underestimate the variability of the βs by as much as 50 percent.

There are other issues as well, particularly if/when there is more direct autoregressiveness (e.g., when the value of Y depends directly on past values); we'll talk about these in a bit.

Solution #1: Fixed Effects

One way of dealing with dependence / heterogeneity in binary-response models is through the use of fixed effects:

$$Y_{it} = f(\mathbf{X}_{it}\beta + \alpha_i + u_{it}) \tag{16}$$

Note that the fixed-effects approach doesn't lend itself to a probit model, so we'll limit the discussion to fixed-effects logit:

$$\Pr(Y_{it} = 1) = \frac{\exp(\mathbf{X}_{it}\beta + \alpha_i)}{1 + \exp(\mathbf{X}_{it}\beta + \alpha_i)} \equiv \Lambda(\mathbf{X}_{it}\beta + \alpha_i) \tag{17}$$

General Issues

The fixed-effects logit specification is both similar to and a bit different than that for continuous dependent variables.

Variances

Because the variance of the latent variable is "lost", we can identify β only up to a scale factor (that is, we can estimate $\frac{\beta}{\sigma_u}$). We typically have to impose some sort of normalization on the variance term – e.g., setting it to 1.0 in the probit case. In the fixed-effects case, we have the conditional variance, $\text{Var}(u_{it}|\alpha_i)$ which is subject to this normalization.

Incidental Parameters and Consistency

As in the case of a continuous variable, we have an incidental parameters problem: as $N \to \infty$, the number of αs increases as well. This means that the fixed-effects logit model is only consistent in T.

There's another problem here, too, however, which is that, unlike in the linear case, *the fact that the αs are inconsistent "pollutes" the estimates of β as well.* This means that, even as $N \to \infty$, we still don't get $\hat{\beta} \to \beta$. Statistically, we can't "sweep out" the heterogeneity by taking deviations from within-group means, the way we did in the linear model. The practical importance of this fact is that, rather than simply estimating a fixed-effects logit model, we need to reconsider how we estimate this model.

Andersen (1970, 1973; see also Chamberlain 1980) came up with a way to rid ourselves of the heterogeneity, by considering a function of β that is independent of the αs. The unconditional likelihood for the model in (17) is similar to that in (6):

$$L^U = \prod_{i=1}^{N}\prod_{t=1}^{T} \Lambda(\mathbf{X}_{it} + \alpha_i)^{Y_{it}}[1 - \Lambda(\mathbf{X}_{it} + \alpha_i)]^{1-Y_{it}} \tag{18}$$

Chamberlain suggested conditioning each set of observations on the number of "1s" for that particular unit. This yields the *conditional likelihood*:

$$L^C = \prod_{i=1}^{N} \Pr\left(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, ...Y_{iT} = y_{iT} \mid \sum_{t=1}^{T} Y_{it}\right) \tag{19}$$

So, for example, suppose that we have a large N and $T = 2$. There are thus four possibilities: $(Y_{i1}, Y_{i2}) = (0,0), (0,1), (1,0)$ and $(1,1)$. Note that:

- $\Pr(Y_{i1} = 0 \text{ and } Y_{i2} = 0 \mid \sum_T Y_{it} = 0) = 1.0$
- $\Pr(Y_{i1} = 1 \text{ and } Y_{i2} = 1 \mid \sum_T Y_{it} = 2) = 1.0$

This means (in the context of unit-level fixed effects) that these contribute a value of 1.0 to the likelihood in (19), and so tell us nothing about the process driving the difference between 0s and 1s. These observations thus "drop out" of the estimation.[1] For the other two possibilities, we have:

[1] Another way to think about this is that when, for some observation i, Y_{it} always equals one, it becomes impossible to estimate α_i, since in order for $\Pr(Y_{it} = 1)$, α_i must equal infinity. Conversely, if Y_{it} always

$$\Pr\left(Y_{i1} = 0 \text{ and } Y_{i2} = 1 \mid \sum_T Y_{it} = 1\right) = \frac{\Pr(0,1)}{\Pr(0,1) + \Pr(1,0)} \tag{20}$$

with a similar statement for $\Pr(Y_{i1} = 0 \text{ and } Y_{i2} = 1 \mid \sum_T Y_{it} = 1)$. Equation (20) yields probabilities in which the αs drop out of the likelihood (see Greene 1997 for a detailed exposition); these can then be maximized. Hence, the same "conditional likelihood", or "conditional fixed effects logit".

The intuition of all this is that, if we know the total number of "1s", we can condition on this to get estimates of β. For a particular unit i, the α_i determines the overall proportion of 1s in the data, while the $\mathbf{X}_i$s (and the βs) determine *when* those 1s occur.

Other Practical Matters

- As with standard fixed effects, we cannot use fixed-effects logit to estimate the effects of variables that don't vary over time.
- Interpretation is standard, but conditional on the value of the fixed effects. That means that, when you generate (e.g.) predicted probabilities, you have to select a value for α_i, since the nonlinearity of the functional form has an impact on the "'location" along that S-curve (and thus on the marginal influence of $\mathbf{X}$ on Y).
- As always, *think* about whether fixed effects make sense; in many instances (e.g., see below), they don't. This is especially true (as in the linear case) when you have very large numbers of (and/or) uninteresting units (e.g., panel surveys, or dyadic IR data).

Fixed Effects and Dyadic IR Data

Green et al. (2001) make a case for fixed effects in the context of studies of international conflict. Their point is that dyadic data almost certainly have heterogeneity (i.e., unobserved, unmeasured variation) by dyad, that if we ignore that heterogeneity we get lousy estimates of the βs, and that fixed effects are a good way of dealing with the heterogeneity. They also show that, if one does account for fixed effects, a big chunk of the "liberal peace" result goes away (e.g., democracy and trade no longer cause peace...).

There are a number of critiques of their findings in the paper, but the ones we care about are those of Beck & Katz, and of King. B&K have several problems with the Green et al. analysis...

- Fixed effects eat up LOTS of degrees of freedom.
- They prevent you from estimating the effects of variables that don't vary over time (e.g., contiguity).
- They don't use any information from dyads that never (or always) go to war ... they assume that the (non-)presence of war in these cases is due entirely to the "proper noun" aspects of the dyad.
- Fixed effects don't make a lot of sense in the context of dyadic IR data; we'd much prefer to have more theoretical explanations for conflict...

I'm with B&K on this one... But, we'll talk a *lot* more about this when we get to survival models.

---

equals zero, then α_i must equal negative infinity. Thus, we can't use these observations in our estimation. In fact, if you try to estimate such a model on such data in Stata, it will drop the observations that are uniformly $Y = 1$ (and/or $Y = 0$), albeit while informing you that it is doing so.

Random-Effects Probit

Random effects models for binary data offer many of the same advantages and disadvantages as do those for continuous data. We'll discuss these, and work through an example or two.

Derivation and Estimation

We start with the same model as before:

$$\begin{aligned} Y_{it}^* &= \mathbf{X}_{it}\beta + u_{it} \\ Y_{it} &= 0 \text{ if } Y_{it}^* \leq 0 \text{ ;} \\ &= 1 \text{ if } Y_{it}^* > 0 \end{aligned}$$

As in the linear case, we can then decompose the error term into two components:

$$u_{it} = \alpha_i + \eta_{it} \tag{21}$$

where we assume that $\eta_{it} \sim$ i.i.d. N(0,1), and that the αs are independent random draws from a normal distribution $\alpha_i \sim N(0, \sigma_\alpha^2)$. This means that:

$$\mathrm{Var}(u_{it}) = 1 + \sigma_\alpha^2 \tag{22}$$

Furthermore, the common error component α_i means that, within units, the u_{it}s will be (equi)correlated. The magnitude of that correlation is given by:

$$\mathrm{Corr}(u_{it}, u_{is},\ t \neq s) \equiv \rho = \frac{\sigma_\alpha^2}{1 + \sigma_\alpha^2} \tag{23}$$

(Note that this, in turn, means that we can write $\sigma_\alpha^2 = \frac{\rho}{1-\rho}$).

Consider first random-effects probit. If the various realizations of Y_{it} for each i were independent, we could simply run a plain-vanilla probit. Because they are correlated, however, things are more difficult; the common α_is mean that the T_i observations on unit i are conditionally distributed according to a T-variate normal distribution. This means that the likelihood is really complicated; the contribution of unit i to the likelihood is:

$$\begin{aligned} L_i &= \mathrm{Prob}(Y_{i1} = y_{i1},\ Y_{i2} = y_{i2},\ ...Y_{iT} = y_{iT}) \\ &= \int_{-\infty}^{X_{i1}\beta} \int_{-\infty}^{X_{i2}\beta} ... \int_{-\infty}^{X_{iT}\beta} \phi(u_{i1}, u_{i2}...u_{iT}) du_{iT}...du_{i2}\ du_{i1} \end{aligned} \tag{24}$$

Similarly, if we consider random-effects logit, we get:

$$\begin{aligned} L_i &= \mathrm{Prob}(Y_{i1} = y_{i1},\ Y_{i2} = y_{i2},\ ...Y_{iT} = y_{iT}) \\ &= \int_{-\infty}^{X_{i1}\beta} \int_{-\infty}^{X_{i2}\beta} ... \int_{-\infty}^{X_{iT}\beta} \lambda(u_{i1}, u_{i2}...u_{iT}) du_{iT}...du_{i2}\ du_{i1} \end{aligned} \tag{25}$$

Obviously, as T gets larger, this problem gets harder and harder. There are two ways of dealing with this issue. The classic econometric approach is due to Butler and Moffitt (1982; and others as well), who figured out that, because the dependence in the us is completely due to the common variation in the αs, we can eliminate the higher-order integrals by conditioning on the αs, and integrating them out of the likelihood. So, for (e.g.) the random–effects probit, this gives:

$$\phi(u_{i1}, u_{i2}, ...u_{iT}) = \int_{-\infty}^{\infty} \phi(u_{i1}, u_{i2}, ...u_{iT} \mid \alpha_i)\phi(\alpha_i)d\alpha_i \tag{26}$$

This approach limits us to evaluating one-dimensional integrals, which is easier. An approximation known as Gauss-Hermite quadrature is most often used to evaluate the likelihood.[2]

A second alternative is to adopt a formally Bayesian approach, and to sample from the posterior of (24) to arrive at estimates. I am somewhat (but not completely) certain that one can do this in WinBUGS, and/or in other Bayesian/MCMC-type statistical packages.

More Practical Stuff

As with the linear case, the standard random-effects probit model makes a few key assumptions, and has a few key characteristics as well:

- Most important is that $\text{Cov}(X_{it}, \alpha_i) = 0$; this assumption is critical in order to get consistent estimates of the βs (but we'll discuss a model that doesn't require this assumption in a minute...).
- The Butler and Moffitt simplification also requires that correlations be equal across all T periods (that is, that the induced dependence in u_{it} is the same, irrespective of the lag at which that dependence occurs). But this is a relatively minor thing, given how much simpler their approach makes computation.
- This model is most appropriate when you have a relatively large N, and relatively small T (that is, with "panel data"). The complexity of the problem (and the inaccuracy of the conditional estimator) grows as T gets larger, and becomes really intractable above $T = 15$ or so (or so says Greene the non-Bayesian...).
- The estimate $\hat{\rho}$ can be thought of as the proportion of the variance due to the random effects, and so tells you something about the importance of unit-specific effects on Y. In point of fact, there are lots of different ways of going about estimating $\hat{\rho}$ – more on this in a minute.
- Stata implements random-effects probit as `-xtprobit, ... re-`. It uses Gauss-Hermite quadrature, and can be pretty slow. Also, in order to ensure that your results are just do to a bad approximation, its a good idea to use `-quadchk-` after estimating such a model; that retries the model using a different number of quadrature "support points", and reports those results. If the results vary a lot by the number of points used, then `-xtprobit-` using quadrature may not be reliable.

Chamberlain's CRE Estimator

As in the linear model, the requirement that $\text{Cov}(\mathbf{X}_{it}, \alpha_i) = 0$ is a doozy. Chamberlain (1984) derives a model in which this assumption is relaxed. He does this by assuming that:

$$\alpha_i = \sum_{t=1}^{T} \mathbf{X}_{it} A_t + \nu_i \tag{27}$$

[2] Gauss-Hermite quadrature is a means of approximating a hard (i.e. non-closed-form) integral. Formally, $\int_L^U W(x)f(x)dx \approx \sum_{j=1}^{M} w_j f(a_j)$, where $W(x)$ is a "weighting function", the $w_j s$ are the "quadrature weights," and the a_js are the "quadrature abcissas" (or "support points") (Greene 1997, 190; Jäckel 2005). The intuition is to approximate the shape of the function by considering a weighted sum of its values at a number of points along its length. The greater the number of support points, the more accurate the approximation.

where $\nu_i \sim N(0, \sigma^2_\nu)$ and $\text{Cov}(\mathbf{X}_{it}, \nu_i) = 0$. That is, the estimator assumes that the relationship between the $\mathbf{X}_{it}$s and the α_is is completely captured by a model that includes the leads and lags of the $\mathbf{X}$s. The basic equation in (14) then becomes:

$$Y^*_{it} = \mathbf{X}_{it}\beta + \sum_{t=1}^{T} \mathbf{X}_{it} A_t + \epsilon_{it} \tag{28}$$

where $\epsilon_{it} = \nu_i + \eta_{it}$.

This general model is known as the "correlated random effects" (CRE) model. This model was introduced to political science by Wawro (2001), in his study of campaign contributions and roll call votes. Note several things about the CRE model:

- Estimation is done by first estimating separate independent probits for each time period, and then "stacking" those reduced-form estimates and using (e.g.) a minimum-distance estimator to impose restrictions on that vector to obtain estimates of β and A_t. Hsiao (1986, 165-7) has details on this; there is also a source for GAUSS code listed in Greg's article.
- An advantage of this approach is that one can do a standard Wald-type test to determine whether or not the Xs and the αs are correlated.
- A disadvantage is that, because we're including leads and lags of $\mathbf{X}_{it}$ in the model, we can no longer include $\mathbf{X}$s that don't vary over time (to do so would induce perfect collinearity). So, that's a potentially large disadvantage, particularly if, as may be the case, some of those variables are correlated with the αs...
- Caveat emptor: some of the assumptions/restrictions necessary to identify this model may be no more realistic than those for a plain-vanilla-random effects probit...

Binary Panel/TSCS Models in Stata

There are a number of these models that can be estimated using Stata. The relevant ones for today are:

- `xtlogit`
  - This is the command to do panel/TSCS binary logistic regression.
  - Options include `fe` (fixed effects), `re` (random effects), and `pa` (population averaged / GEE).
  - Random effects models are estimated via Gauss-Hermite quadrature; use the `intp` option to denote the number of quadrature points (default = 12). You can also specify adaptive or nonadaptive Gauss-Hermite quadrature; I don't recommend mucking around with this unless you have to...
- `xtprobit`
  - This is (surprise!) the command to estimate panel/TSCS binary probit models.
  - Options are the same as for `xtlogit`, but without fixed effects: `re` is for random effects, and `pa` gets you population averaged / GEE models.
  - Same holds for the `intp` option for `re` models.
- `xtcloglog`
  - Estimates random-effects (`re` option) and population-averaged (`pa` option) models with a complimentary log-log link: $\Pr(Y_{it} = 1) = 1 - \exp[-\exp(\mathbf{X}_{it}\beta)]$. This is a non-symmetrical CDF corresponding to an extreme-value PDF.

There are also a few useful non-estimation commands for these sort of data...

- `xttrans`
  - A command that displays *transition probabilities* – that is, a unit–specific crosstab of Y_{it} and Y_{it-1}.
  - The `freq` option also causes frequencies to be displayed; otherwise, the table just contains percentages.
  - This can be useful for getting a sense of whether or not there are temporal trends in the data of any importance.

- `quadchk`
  - Used after any of the `re` models mentioned above.
  - A robustness check command; it reestimates the immediately preceding random effects model, using both smaller (-4) and larger (+4) numbers of quadrature points; it then presents the absolute and relative differences in the coefficients across the three models.
  - In general, we'd like the relative differences to be less than one percent (i.e., 0.01) to be sure that the (arbitrary) choice of quadrature points isn't driving the results.

- `xtrho` / `xtrhoi`
  - Estimates the within-unit covariance following any of the random effects models described above.
  - According to its help file, "The distribution of the possible outcomes for two observations in the same group (as defined by the `i()` option in the estimation command) may be viewed as a 2 × 2 contingency table, with cell probabilities depending on the values of the linear predictor ($\mathbf{X}\hat{\beta}$) and the standard deviation of the random effect (σ_α). `xtrho` estimates this table using numerical integration and reports the marginal probability of a positive outcome and the joint probability of two positive outcomes in the same group. Together, these two quantities define all four cell probabilities. The command then calculates three standard measures of association based on the 2 × 2 table: Pearson's linear correlation coefficient r, the odds ratio, and Yule's Q coefficient, which in 2 × 2 tables coincides with Goodman and Kruskal's γ."
  - The command also reports the predicted marginal probability of a positive outcome (that is, $\Pr(Y_{it} = 1 \mid \mathbf{X}\hat{\beta}, \sigma_\alpha)$), as well as the predicted joint probability (that is, $\Pr(Y_{it} = 1 \,\text{and}\, Y_{is} = 1 \mid \mathbf{X}\hat{\beta}, \sigma_\alpha)$).
  - It does this for $\mathbf{X}\hat{\beta}$ values equal to the data medians in the estimation sample, and with $\sigma_\alpha = \hat{\sigma}_\alpha$, and reports standard errors for all quantities of interest.
  - `xtrhoi` allows the user to set $\mathbf{X}\hat{\beta}$ to some value(s) of particular interest.
  - The command also contains a very useful detail option that will calculate the same quantities for the first, 25th, 50th, 75th, and 99th percentiles of the data.

An Example: Comparing Fixed and Random Effects for Binary Data

To compare fixed and random effects models for binary data, I look at data from Segal (1986), on Supreme Court search and seizure cases, 1962-1981. The dependent variable is whether each justice ($N = 14$) voted to allow the search to occur (=1) or not (=0) in each case ($\bar{T} = 74.1$). This is a function of eight variables:

- `warrant`: Whether (=1) or not (=0) a warrant was issued,
- `house`: Whether (=1) or not (=0) the search was of a private home,

- `person`: Whether (=1) or not (=0) the search was of a person,
- `business`: Whether (=1) or not (=0) the search was of a business,
- `car`: Whether (=1) or not (=0) the search was of an automobile,
- `us`: Whether (=1) or not (=0) the U.S. government was the petitioner,
- `except`: The number of "exceptions" outlined by the Court under which the search fell, and
- `justideo`: The justice's Segal-Cover (1989) ideology score, ranging from zero (most conservative) to 1 (most liberal).

Note that:

1. Because the justices' Segal-Cover scores don't vary over "time" (that is, cases), the fixed effects model can't estimate the coefficient for that variable.
2. The results are generally pretty similar, though the coefficients are of different sizes (because one is logit, the other probit).
3. The `-quadchk-` indicates that there isn't a large amount of variation in the coefficients as a function of the number of quadrature points used. (The rule of thumb is that you want less than a 1% relative change from one set of points to another). This is good; it means that we can be relatively confident that our results aren't driven by that (mostly arbitrary) decision.

Tobit Models for Censored Panel Data

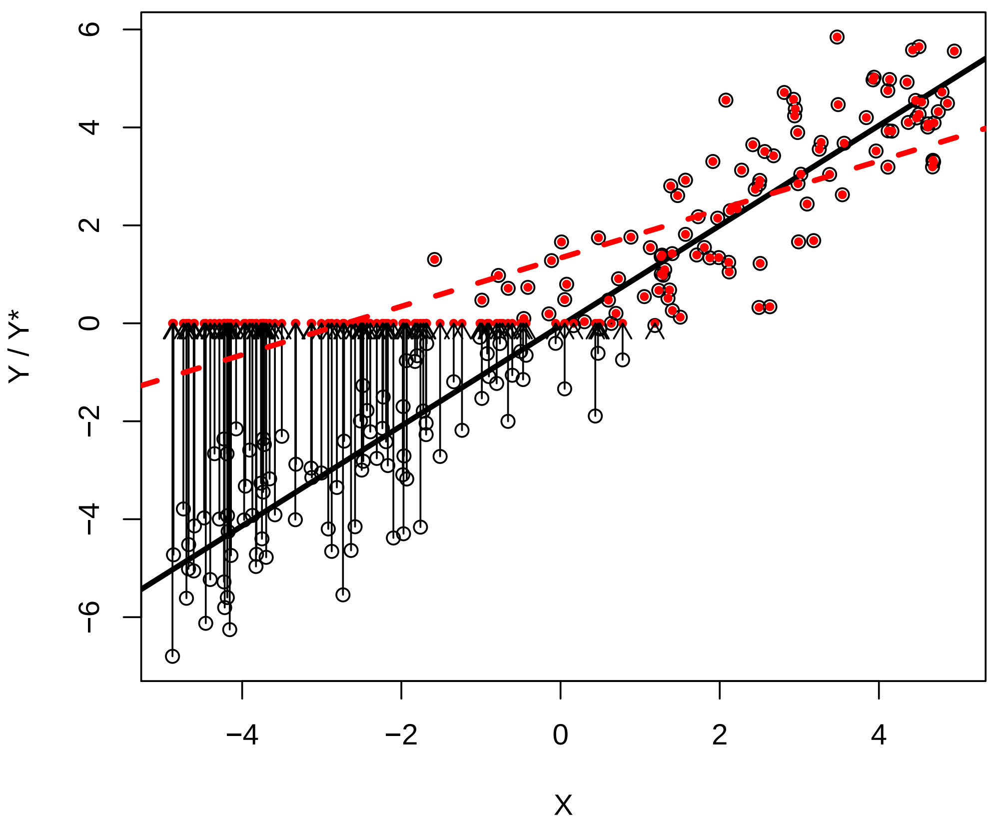
Another form of data we often encounter is *censored* data, of the form:

$$\begin{aligned} Y_i &= Y_i^* \text{ if } Y_i^* > L \\ &= L \text{ if } Y_i^* \leq L \end{aligned}$$

Here, Y is censored from below; it is also possible to have "top"-censoring, or censoring both from below and from above. Often, censoring from below occurs at a value of zero, as in cases where the response measures the extent of some trait or characteristic and negative values are not possible.

It's long been well-understood that using standard methods (e.g., OLS) on censored data will yield estimates that are both biased and inconsistent (e.g., Tobin 1958; Amemiya 1973). The direction of the bias is towards zero; the magnitude of the bias is more-or-less proportional to the fraction of data that are censored (e.g., Goldberger 1972; Greene 1981). Intuitively, the reason for this is that the censoring mechanism nonrandomly attenuates the variability in Y, leading to incorrect estimates of β. Additionally, the estimator is inconsistent because (intuitively) no amount of additional (censored) data will cause the bias to disappear (see Figure 2).

Figure 2: Tobit – Intuition



Suppose (as we've been doing in the linear case) we begin with a model like

$$Y_i^* = \mathbf{X}_i\boldsymbol{\beta} + u_i, \tag{29}$$

and the usual assumption that $u_i \sim$ i.i.d. $N(0, \sigma^2$. The observed variable Y_i is therefore something like a mix between a binary indicator (below the cut-point L) and a more-or-less standard linear regression model above L. In particular, we can think of two types of observations: those with $Y^* > L$, and those with $Y^* \leq L$.

For the former cases, we observe the "real" value of Y^*, such that the likelihood for those observations conforms to a standard linear model:

$$\mathbf{L}_1(\beta, \sigma^2|Y, L) = \prod_{Y_i > L} \phi(Y_i^*|\mathbf{X}_i, \beta, \sigma^2). \tag{30}$$

For the cases with $Y = L$, we know only that $Y^* \leq L$; we can think of this as one outcome of a binary response (i.e., $Y = L$ vs. $Y \neq L$). This suggests that:

$$\begin{aligned}
\Pr(Y_i = L) &= \Pr(Y_i^* \leq L) \\
&= \int_{-\infty}^{L} \phi(Y_i^*|\mathbf{X}_i, \beta, \sigma^2)\, dY^* \\
&= \Phi(L|\mathbf{X}_i, \beta, \sigma^2). \qquad (31)
\end{aligned}$$

This second term is a lot like a probit probability, and leads to a likelihood for the observations with $Y = L$ of:

$$\mathbf{L}_2(\beta, \sigma^2|Y, L) = \prod_{Y_i = L} \Phi(L|\mathbf{X}_i, \beta, \sigma^2). \tag{32}$$

The combined likelihood for the two models is then equal to

$$\mathbf{L}(\beta, \sigma^2|Y, L) = \prod_{Y_i > L} \phi(Y_i^*|\mathbf{X}_i, \beta, \sigma^2) \prod_{Y_i = L} \Phi(L|\mathbf{X}_i, \beta, \sigma^2). \tag{33}$$

This is the tobit model (for "Tobin's probit"). Amemiya (1973) showed that the likelihood in (33) yields consistent estimates of $\hat{\beta}$ and $\hat{\sigma}^2$. In addition, the βs can be interpreted in the standard way (i.e., as if they were OLS estimates).

A Panel Tobit Estimator

A number of individuals have developed modified Tobit estimators for panel/TSCS data. The two most commonly-seen are fixed- and random-effects models for unit intercepts:

$$Y_i^* = \mathbf{X}_i\boldsymbol{\beta} + \alpha_i + u_i \tag{34}$$

As with the logit model we discussed yesterday, the unconditional fixed effects Tobit model (that is, estimating the model in (33) with $N - 1$ indicator variables for each unit) yields coefficient estimates that are biased, for the same intuitive reasons as we discussed in the logit case. However, unlike the logit case (where we could condition on the sufficient statistic $\sum_T Y_{it}$ to remove the unit effects and derive a conditional estimator), there is no analogous sufficient statistic in the Tobit case. As a result, fixed effects are generally not estimated in a Tobit context.[3]

More commonly used is a random-effects specification of (34); this model is available in `Stata` as `xttobit`, and can also be estimated in `R` using the parametric models available in the `survival` package. We'll review an illustration of this model in the examples.

[3]Honore (1992 *Econometrica*) develops a semiparametric fixed-effects Tobit estimator; read his article if this sounds like something you might have a use for.

Models for Panel/TSCS Event Counts

There are basically three kinds of panel/TSCS models for event count data: fixed-effects, random-effects and marginal (i.e., GEE) models.

Event Count Data

Data where the response variable is a nonnegative integer count – the number of occurrences of some event in a given period (or other defined unit).

The most basic model for counts of events is the Poisson:

$$\Pr(Y_i = y) = \frac{\exp(-\lambda_i)\lambda_i^y}{y!} \tag{35}$$

where μ_i is the expected value of Y_i and we typically introduce covariates as $\mu_i = \exp(\mathbf{X}_i\beta)$. I won't go into the Poisson distribution that much here; take the MLE class for the details. Importantly, though, the Poisson distribution has the property that $\mathrm{E}(Y) = \mathrm{Var}(Y)$; if this is not the case, then other models (e.g., the negative binomial model) are necessary.

Fixed Effects Models for Count Data

The fixed-effects Poisson model is:

$$Y_{it} \sim \text{Poisson}(\mu_{it} = \alpha_i \lambda_{it}) \tag{36}$$

where, as above, $\lambda_{it} = \exp(\mathbf{X}_{it}\beta)$. Note that this specification means that the conditional expectation of Y is:

$$\begin{aligned} \mathrm{E}(Y_{it} \mid \mathbf{X}_{it}, \alpha_i) &= \mu_{it} \\ &= \alpha_i \exp(\mathbf{X}_{it}\beta) \\ &= \exp(\delta_i + \mathbf{X}_{it}\beta) \end{aligned} \tag{37}$$

where $\delta_i = \ln(\alpha_i)$.

If N is small (and fixed), this model can be estimated in the standard fashion, simply by including dummy variables for each of the N units i. Thus, as with other kinds of models, fixed-effects are reasonable only when N is small relative to T, and when N is not increasing (that is, when one's asymptotics don't need to depend on N). However, Lancaster (in Cameron and Trivedi 1998) shows that there is no "incidental parameters problem" for Poisson regression, using a "concentrated likelihood" (see Cameron and Trivedi, pp. 281-2). One therefore needs to use a "conditional" approach, similar to that for fixed-effects logit, in which the fixed effects are conditioned on the sum of the event counts within the panel and then concentrated out of the likelihood. As with conditional fixed-effects logit, this relies on the fact that observations across units are independent (i.e., that the fixed effects completely capture the unobserved heterogeneity in the data).

In `Stata`, this is accomplished through the `-xtpois, fe-` command. Interpretation is standard for Poisson models, with the usual caveat that results are conditional on the values of the fixed effects. Note as well that, for overdispersed data, there is also a conditional fixed-effects negative binomial model (in `Stata`, `-xtnbreg, fe-`). In `R`, we can estimate the same model using the `glmmML` package.

Random-Effects Models for Event Count Data

The random-effects Poisson model is similar to that for (e.g.) logit and probit, in that we assume that the α_is are distributed as some i.i.d. random variable, which we then integrate out of the likelihood:

$$\begin{aligned}\Pr(Y_{i1} = y_{i1}, ...Y_{iT} = y_{iT}) &= \int_0^\infty \Pr(Y_{i1} = y_{i1}, ...Y_{iT} = y_{iT}) f(\alpha_i) d\alpha_i \\ &= \int_0^\infty \left[\prod_{t=1}^{T} \Pr(Y_{it} \mid \alpha_i)\right] f(\alpha_i) d\alpha_i \qquad (38)\end{aligned}$$

The simplest distribution to use for the α_is is the *Gamma*, with an expected value and variance equal to some parameter θ. (The gamma is the conjugate distribution for the Poisson.) The result is a likelihood (see Cameron and Trivedi, p. 288) which is relatively straightforward to maximize. The random-effects Poisson model has $\mathrm{E}(Y_{it}) = \lambda_{it}$ and $\mathrm{Var}(Y_{it}) = \lambda_{it} + \frac{\lambda_{it}^2}{\theta}$. The `Stata` command is `-xtpois, re-`, while `glmmML` is the way to go in `R`; and interpretation is also as with other conditional-effect models.

One can also estimate this model assuming that the α_is are distributed normally, in which case the quadrature procedure used for random-effects probit (or some other approach, e.g. MCMC, a la Chib et al. 1998) is necessary. Note as well that there is also a random-effects negative binomial model due to Hausman, Hall and Griliches (1984), which is available only for gamma-distributed α_is and which is estimable in `Stata` using `-xtnbreg, re-`; negative binomial models are not included in `glmmML`.

An Example

Our example examines data from Phase III of the State Failure Task Force report, and includes examples of panel-data Tobit and event count models; see the slides for details.

Appendix: Materials from the Old Handout

State Failure Task Force Data

```
. xtdes

countryid:  2, 3, ..., 195                                 n =        170
    year:  1957, 1962, ..., 1997                           T =          9
           Delta(year) = 5; (1997-1957)/5 + 1 = 9
           (countryid*year uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       2       6         8         9       9       9

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+-----------
       79     46.47   46.47 |  111111111
       28     16.47   62.94 |  .11111111
       22     12.94   75.88 |  .......11
        9      5.29   81.18 |  ...111111
        9      5.29   86.47 |  ..1111111
        7      4.12   90.59 |  ....11111
        5      2.94   93.53 |  ........1
        5      2.94   96.47 |  1111111..
        2      1.18   97.65 |  1111.....
        4      2.35  100.00 | (other patterns)
 ---------------------------+-----------
      170    100.00         |  XXXXXXXXX


. su countryid year  SumEvents ciob cioc POLITY unuurbpc poldurab

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   countryid |      1203     97.1064    54.72243          2        195
        year |      1203    1978.779    12.58559       1957       1997
   SumEvents |      1194    5.842127    11.88424          0         61
        ciob |      1203    18.78554     8.11198          0         38
        cioc |      1203    5.595179    4.541347          0         24
-------------+--------------------------------------------------------
      POLITY |      1189   -.6644239    7.531924        -10         10
    unuurbpc |      1146     43.3468    24.57294       2.03        100
    poldurab |      1198    20.91653    22.83868          0         97
```

Panel Tobit Models

```
. xtreg SumEvents POLITY unuurbpc poldurab year, re

Random-effects GLS regression                   Number of obs      =      1132
Group variable (i): countryid                   Number of groups   =       160
R-sq:  within  = 0.1273                         Obs per group: min =         1
       between = 0.0812                                        avg =       7.1
       overall = 0.0841                                        max =         9
Random effects u_i ~ Gaussian                   Wald chi2(4)       =    140.29
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
------------------------------------------------------------------------------
   SumEvents |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      POLITY |  -.1771263   .0554206    -3.20   0.001    -.2857486   -.0685039
    unuurbpc |  -.0792978   .0274317    -2.89   0.004     -.133063   -.0255326
    poldurab |  -.1275142   .0201363    -6.33   0.000    -.1669806   -.0880478
        year |   .2667588    .024654    10.82   0.000     .2184379    .3150797
       _cons |    -516.75   48.19351   -10.72   0.000    -611.2075   -422.2924
-------------+----------------------------------------------------------------
     sigma_u |  8.3658691
     sigma_e |  7.7868834
         rho |   .5357984   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. xttobit SumEvents POLITY unuurbpc poldurab year, re ll(0)

Random-effects tobit regression                 Number of obs      =      1132
Group variable (i): countryid                   Number of groups   =       160
Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       7.1
                                                               max =         9
                                                Wald chi2(4)       =    196.09
Log likelihood  = -2042.2823                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
   SumEvents |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      POLITY |  -.4908305   .1005571    -4.88   0.000    -.6879188   -.2937422
    unuurbpc |  -.1194613   .0505678    -2.36   0.018    -.2185723   -.0203503
    poldurab |  -.2816857   .0386693    -7.28   0.000    -.3574762   -.2058952
        year |    .552916   .0500738    11.04   0.000     .4547731    .6510588
       _cons |  -1085.844   98.46183   -11.03   0.000    -1278.826   -892.8623
-------------+----------------------------------------------------------------
    /sigma_u |   18.94774   1.593354    11.89   0.000     15.82482    22.07066
    /sigma_e |   12.66688   .4636415    27.32   0.000     11.75816     13.5756
-------------+----------------------------------------------------------------
         rho |   .6911259   .0384133                      .6122919    .7620124
------------------------------------------------------------------------------
  Observation summary:       707  left-censored observations
                             425     uncensored observations
                               0 right-censored observations
```

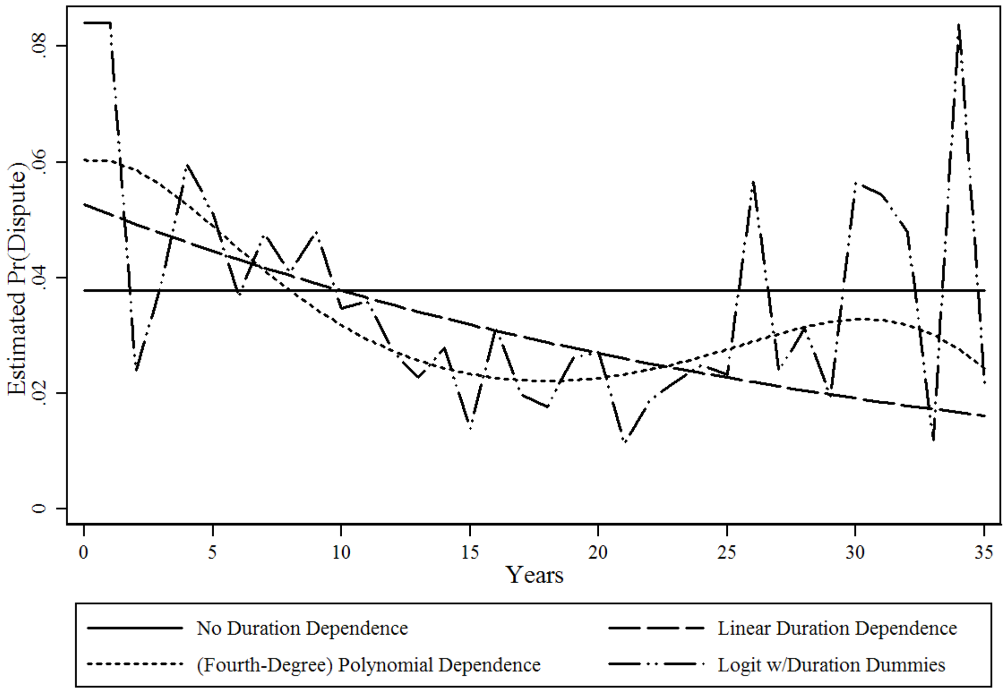
Panel Models for Event Counts

```
. poisson ciob POLITY unuurbpc poldurab year

Poisson regression                                Number of obs   =       1132
                                                  LR chi2(4)      =    1547.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -4059.5127                       Pseudo R2       =     0.1601
------------------------------------------------------------------------------
        ciob |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      POLITY |   .0103559   .0009817    10.55   0.000     .0084317      .01228
    unuurbpc |   .0048643   .0003201    15.20   0.000      .004237    .0054916
    poldurab |   .0020246   .0002948     6.87   0.000     .0014468    .0026025
        year |    .011826    .000569    20.78   0.000     .0107107    .0129412
       _cons |  -20.74132   1.125035   -18.44   0.000    -22.94635   -18.53629
------------------------------------------------------------------------------


. xtpoisson  ciob POLITY unuurbpc poldurab year, fe
note: 5 groups (5 obs) dropped because of only one obs per group

Conditional fixed-effects Poisson regression    Number of obs      =      1127
Group variable (i): countryid                   Number of groups   =       155
                                                Obs per group: min =         2
                                                               avg =       7.3
                                                               max =         9
                                                Wald chi2(4)       =   1208.13
Log likelihood  = -2558.3941                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        ciob |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      POLITY |  -.0074369   .0019391    -3.84   0.000    -.0112375   -.0036363
    unuurbpc |   .0050111   .0015795     3.17   0.002     .0019153     .008107
    poldurab |  -.0004774   .0007489    -0.64   0.524    -.0019452    .0009905
        year |   .0184112   .0011151    16.51   0.000     .0162257    .0205966
------------------------------------------------------------------------------


. est store fixed
```

```
. xtpoisson ciob POLITY unuurbpc poldurab year, re

Random-effects Poisson regression               Number of obs      =      1132
Group variable (i): countryid                   Number of groups   =       160
Random effects u_i ~ Gamma                      Obs per group: min =         1
                                                               avg =       7.1
                                                               max =         9
                                                Wald chi2(4)       =   1181.51
Log likelihood  =  -3390.628                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        ciob |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      POLITY |  -.0034547   .0017936    -1.93   0.054    -.0069701    .0000608
    unuurbpc |   .0061469   .0010388     5.92   0.000     .0041109    .0081829
    poldurab |    .001147   .0006596     1.74   0.082    -.0001458    .0024398
        year |   .0163423   .0008313    19.66   0.000     .0147129    .0179717
       _cons |   -29.7832   1.618543   -18.40   0.000    -32.95549   -26.61092
-------------+----------------------------------------------------------------
    /lnalpha |  -1.951291   .1266443                     -2.199509   -1.703072
-------------+----------------------------------------------------------------
       alpha |   .1420906    .017995                      .1108576    .1821231
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0: chibar2(01) =  1337.77 Prob>=chibar2 = 0.000


. hausman fixed .

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
      POLITY |   -.0074369    -.0034547       -.0039822         .000737
    unuurbpc |    .0050111     .0061469       -.0011358        .0011899
    poldurab |   -.0004774      .001147       -.0016244        .0003546
        year |    .0184112     .0163423        .0020689        .0007431
------------------------------------------------------------------------------
                     b = consistent under Ho and Ha; obtained from xtpoisson
        B = inconsistent under Ha, efficient under Ho; obtained from xtpoisson

    Test:  Ho:  difference in coefficients not systematic

                  chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      293.49
                Prob>chi2 =      0.0000
```

Appendix: Material from Old Handout

Figure 3: Probit, Logit, and C-Log-Log CDFs



Segal (1986) Search and Seizure Voting Data...

```
. su vote warrant house person business car us except justideo

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        vote |      1037    .5255545    .4995875          0          1
     warrant |      1037    .1456123    .3528873          0          1
       house |      1037    .2266152    .4188436          0          1
      person |      1037    .3114754    .4633201          0          1
    business |      1037    .1523626    .3595454          0          1
-------------+--------------------------------------------------------
         car |      1037    .1996143    .3999033          0          1
          us |      1037    .4541948    .4981377          0          1
      except |      1037    .3539055    .6033343          0          3
    justideo |      1037    .5897155     .351096       .045          1
```

```
. logit vote warrant house person business car us except justideo
```

```
Logistic regression                               Number of obs   =       1037
                                                  LR chi2(8)      =     238.15
                                                  Prob > chi2     =     0.0000
Log likelihood = -598.36149                       Pseudo R2       =     0.1660

------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     warrant |   .5335293   .2083363     2.56   0.010     .1251976    .9418609
       house |  -1.083996    .275565    -3.93   0.000    -1.624094   -.5438991
      person |  -.9438459    .256926    -3.67   0.000    -1.447412   -.4402801
    business |  -1.472197   .2974792    -4.95   0.000    -2.055245   -.8891481
         car |  -1.006622   .2816294    -3.57   0.000    -1.558606   -.4546385
          us |    .482438   .1482433     3.25   0.001     .1918864    .7729896
      except |    .864028    .138396     6.24   0.000     .5927768    1.135279
    justideo |  -2.402596   .2157913   -11.13   0.000    -2.825539   -1.979653
       _cons |   1.941878   .2798735     6.94   0.000     1.393336     2.49042
------------------------------------------------------------------------------
```

```
. xtlogit vote warrant house person business car us except justideo, fe
note: multiple positive outcomes within groups encountered.
note: justideo omitted due to no within-group variance.

Conditional fixed-effects logistic regression   Number of obs      =      1037
Group variable (i): justid                      Number of groups   =        14
                                                Obs per group: min =        24
                                                               avg =      74.1
                                                               max =       121
                                                LR chi2(7)         =    110.95
Log likelihood  = -494.48223                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     warrant |   .5911754   .2266173     2.61   0.009     .1470136    1.035337
       house |  -1.450236   .3027743    -4.79   0.000    -2.043663   -.8568096
      person |  -1.105403   .2794111    -3.96   0.000    -1.653039   -.5577677
    business |  -1.807048   .3235041    -5.59   0.000    -2.441105   -1.172992
         car |  -1.182927   .3054959    -3.87   0.000    -1.781688   -.5841663
          us |   .5291579   .1606706     3.29   0.001     .2142494    .8440664
      except |    1.07677   .1541784     6.98   0.000     .7745856    1.378954
------------------------------------------------------------------------------

. xtlogit vote warrant house person business car us except justideo, re

Random-effects logistic regression              Number of obs      =      1037
Group variable (i): justid                      Number of groups   =        14
Random effects u_i ~ Gaussian                   Obs per group: min =        24
                                                               avg =      74.1
                                                               max =       121
                                                Wald chi2(8)       =     98.99
Log likelihood  = -547.81799                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     warrant |   .5941799   .2258876     2.63   0.009     .1514483    1.036912
       house |  -1.433851   .3029375    -4.73   0.000    -2.027598   -.8401043
      person |  -1.103904   .2796611    -3.95   0.000     -1.65203   -.5557785
    business |  -1.798843   .3236931    -5.56   0.000     -2.43327   -1.164416
         car |  -1.180339   .3057377    -3.86   0.000    -1.779574   -.5811046
          us |   .5311769   .1604217     3.31   0.001     .2167561    .8455977
      except |   1.070343   .1539737     6.95   0.000     .7685604    1.372126
    justideo |   -2.34464   .7381789    -3.18   0.001    -3.791444   -.8978355
       _cons |   2.016532    .565363     3.57   0.000     .9084412    3.124623
-------------+----------------------------------------------------------------
    /lnsig2u |   -.150106   .4286465                     -.9902378    .6900257
-------------+----------------------------------------------------------------
     sigma_u |   .9276943   .1988265                      .6094984    1.412008
         rho |   .2073533   .0704514                      .1014619    .3773476
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =   101.09 Prob >= chibar2 = 0.000
```

```
. quadchk
                       Quadrature check

             Fitted       Comparison     Comparison
             quadrature   quadrature     quadrature
             12 points    8 points       16 points
-----------------------------------------------------
Log          -547.81799   -547.56451     -547.82111
likelihood                   .253479     -.00311279   Difference
                          -.00046271       5.682e-06   Relative difference
-----------------------------------------------------
vote:         .59417991    .59590261      .59333942
 warrant                   .00172269     -.00084049   Difference
                           .00289928     -.00141454   Relative difference
-----------------------------------------------------
vote:        -1.4338509    -1.432193     -1.4371416
   house                   .00165794     -.00329071   Difference
                          -.00115628      .00229502   Relative difference
-----------------------------------------------------
vote:        -1.1039043   -1.1032623     -1.1052357
  person                   .00064197     -.00133141   Difference
                          -.00058155      .00120609   Relative difference
-----------------------------------------------------
vote:        -1.7988429   -1.7987557     -1.8010852
business                   .00008726     -.00224225   Difference
                          -.00004851       .0012465   Relative difference
-----------------------------------------------------
vote:        -1.1803395   -1.1800101     -1.1816055
     car                   .00032935     -.00126608   Difference
                          -.00027903      .00107264   Relative difference
-----------------------------------------------------
vote:         .53117691    .53237692      .53072834
      us                   .00120001     -.00044857   Difference
                           .00225916     -.00084448   Relative difference
-----------------------------------------------------
vote:         1.0703434    1.0708716      1.0712912
  except                   .00052818      .00094781   Difference
                           .00049347      .00088552   Relative difference
-----------------------------------------------------
vote:        -2.3446395   -2.2697539     -2.4082016
justideo                   .07488561     -.06356202   Difference
                          -.03193907      .02710951   Relative difference
-----------------------------------------------------
vote:         2.0165323     2.036018      2.0010609
   _cons                   .01948562     -.01547142   Difference
                           .00966294     -.00767229   Relative difference
-----------------------------------------------------
lnsig2u:     -.15010605   -.18843793     -.05033698
   _cons                  -.03833189      .09976907   Difference
                           .25536537      -.6646572   Relative difference
-----------------------------------------------------
```

Reversing i and t...

```
. tsset caseid justid
       panel variable:  caseid, 1 to 123
        time variable:  justid, 1 to 14, but with gaps

. xtlogit vote justideo, fe
note: multiple positive outcomes within groups encountered.
note: 24 groups (205 obs) dropped due to all positive or
      all negative outcomes.
Conditional fixed-effects logistic regression   Number of obs      =       832
Group variable (i): caseid                      Number of groups   =        99
                                                Obs per group: min =         6
                                                               avg =       8.4
                                                               max =         9
                                                LR chi2(1)         =    244.77
Log likelihood  = -240.63245                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    justideo |  -4.755163   .4001315   -11.88   0.000    -5.539407    -3.97092
------------------------------------------------------------------------------

. xtlogit vote warrant house person business car us except justideo, re

Random-effects logistic regression              Number of obs      =      1037
Group variable (i): caseid                      Number of groups   =       123
Random effects u_i ~ Gaussian                   Obs per group: min =         6
                                                               avg =       8.4
                                                               max =         9
                                                Wald chi2(8)       =    156.93
Log likelihood  =  -505.6138                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     warrant |   .8458856   .5639042     1.50   0.134    -.2593464    1.951117
       house |  -1.384585   .7358594    -1.88   0.060    -2.826843     .057673
      person |  -1.436958   .6908442    -2.08   0.038    -2.790988   -.0829287
    business |  -2.008578   .7849524    -2.56   0.011    -3.547057   -.4700997
         car |  -1.472248   .7601121    -1.94   0.053    -2.962041    .0175439
          us |   .7163863   .4088281     1.75   0.080    -.0849019    1.517675
      except |   1.248149   .3685106     3.39   0.001     .5258815    1.970417
    justideo |  -4.512786   .3731841   -12.09   0.000    -5.244213   -3.781359
       _cons |   3.393873   .6891457     4.92   0.000     2.043173    4.744574
-------------+----------------------------------------------------------------
    /lnsig2u |   1.300159   .2023601                      .9035408    1.696778
-------------+----------------------------------------------------------------
     sigma_u |   1.915693   .1938299                      1.571091     2.33588
         rho |   .5273008   .0504392                      .4286634    .6238519
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =   185.50 Prob >= chibar2 = 0.000
```

```
. xtprobit vote warrant house person business car us except justideo, re

Random-effects probit regression                Number of obs      =      1037
Group variable (i): justid                      Number of groups   =        14
Random effects u_i ~ Gaussian                   Obs per group: min =        24
                                                               avg =      74.1
                                                               max =       121
                                                Wald chi2(8)       =    106.79
Log likelihood  =  -547.7384                    Prob > chi2        =    0.0000
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     warrant |   .3703301   .1345058     2.75   0.006     .1067036    .6339566
       house |  -.8384161   .1772729    -4.73   0.000    -1.185865   -.4909677
      person |  -.6150002   .1622197    -3.79   0.000    -.9329449   -.2970555
    business |  -1.065552   .1876203    -5.68   0.000    -1.433281   -.6978225
         car |  -.6675218   .1777199    -3.76   0.000    -1.015846   -.3191972
          us |    .314654   .0943338     3.34   0.001     .1297631     .499545
      except |   .6349944   .0888851     7.14   0.000     .4607828     .809206
    justideo |  -1.377226   .4349103    -3.17   0.002    -2.229634   -.5248173
       _cons |   1.164314   .3313048     3.51   0.000     .5149686     1.81366
-------------+----------------------------------------------------------------
    /lnsig2u |  -1.198905   .4249006                     -2.031695   -.3661148
-------------+----------------------------------------------------------------
     sigma_u |   .5491123   .1166591                      .3620955    .8327204
         rho |   .2316701   .0756319                      .1159151    .4094802
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =   100.31 Prob >= chibar2 = 0.000

. hausman fixedeffects randomeffects

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |  fixedeffects randomeffects   Difference          S.E.
-------------+----------------------------------------------------------------
     warrant |    .5911754     .5941799       -.0030046        .0181707
       house |   -1.450236    -1.433851       -.0163854               .
      person |   -1.105403    -1.103904        -.001499               .
    business |   -1.807048    -1.798843       -.0082053               .
         car |   -1.182927    -1.180339       -.0025878               .
          us |    .5291579     .5311769        -.002019        .0089383
      except |     1.07677     1.070343        .0064263        .0079414
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtlogit
            B = inconsistent under Ha, efficient under Ho; obtained from xtlogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        2.57
                Prob>chi2 =      0.9215
                (V_b-V_B is not positive definite)
```

Complementary Log-Log Model with Random Effects

```
. xtcloglog vote warrant house person business car us except justideo, re

Random-effects complementary log-log model      Number of obs      =      1037
Group variable (i): justid                      Number of groups   =        14

Random effects u_i ~ Gaussian                   Obs per group: min =        24
                                                               avg =      74.1
                                                               max =       121

                                                Wald chi2(8)       =    107.03
Log likelihood  = -548.09625                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     warrant |    .475615   .1485608     3.20   0.001     .1844413    .7667887
       house |  -.8349422   .1878358    -4.45   0.000    -1.203094   -.4667909
      person |  -.6249166    .166492    -3.75   0.000     -.951235   -.2985983
    business |  -1.172796   .1995928    -5.88   0.000     -1.56399   -.7816008
         car |  -.6642929   .1829155    -3.63   0.000    -1.022801   -.3057852
          us |   .3441369   .1037698     3.32   0.001     .1407518    .5475219
      except |   .7031655   .0960792     7.32   0.000     .5148537    .8914772
    justideo |  -1.525852   .4986901    -3.06   0.002    -2.503267   -.5484372
       _cons |   .7518567   .3668267     2.05   0.040     .0328896    1.470824
-------------+----------------------------------------------------------------
    /lnsig2u |  -.9182567   .4244445                     -1.750153   -.0863608
-------------+----------------------------------------------------------------
     sigma_u |   .6318342   .1340893                      .4168302    .9577386
         rho |   .1952962   .0667038                      .0955348    .3579987
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =    98.06 Prob >= chibar2 = 0.000
```

Binary-Response Panel Models in `R`

```
> library(foreign)

> SegalVotes<-read.dta("SegalVotes.dta")

> library(glmmML)
```

Random-Effects Logit

```
> SegalRE<-glmmML(vote~warrant+house+person+business+car+us+except+justideo,
  data=SegalVotes, family="binomial", cluster=justid)

> summary(SegalRE)

Call:  glmmML(formula = vote ~ warrant + house + person + business + car +
       us + except + justideo, family = "binomial", data = SegalVotes,
       cluster = justid)

               coef se(coef)      z Pr(>|z|)
(Intercept)  2.0164   0.5648  3.570 3.56e-04
warrant      0.5942   0.2259  2.630 8.53e-03
house       -1.4340   0.3029 -4.734 2.21e-06
person      -1.1041   0.2797 -3.948 7.89e-05
business    -1.7991   0.3237 -5.558 2.73e-08
car         -1.1805   0.3058 -3.861 1.13e-04
us           0.5312   0.1604  3.311 9.29e-04
except       1.0704   0.1540  6.952 3.59e-12
justideo    -2.3445   0.7372 -3.180 1.47e-03

Standard deviation in mixing distribution:  0.9263
Std. Error:                                 0.1952

Residual deviance: 1096 on 1027 degrees of freedom      AIC: 1116
```