NONLINEAR MODELS

# Correlated Random Effects Panel Data Models

IZA Summer School in Labor Economics
May 13-19, 2013

Jeffrey M. Wooldridge
Michigan State University

1. Why Nonlinear Models?
2. CRE versus Other Approaches
3. Nonlinear Unobserved Effects Models
4. Assumptions
5. Correlated Random Effects Probit
6. CRE Tobit
7. CRE Count Models
8. Nonparametric and Flexible Parametric Approaches

## 1. Why Nonlinear Models?

- Suppose $y_{it}$ is binary, $\mathbf{x}_{it}$ is a set of observed explanatory variables, $c_i$ is heterogeneity. We are interested in the response probability as a function of $(\mathbf{x}_t, c)$:

$$p(\mathbf{x}_t, c) = P(y_{it} = 1 | \mathbf{x}_{it} = \mathbf{x}_t, c_i = c).$$

- Because $p(\mathbf{x}_t, c)$ is a probability, a linear model, say

$$p(\mathbf{x}_t, c) = \mathbf{x}_t \boldsymbol{\beta} + c,$$

can be a poor approximation.

- Or, suppose $y_{it} \geq 0$. An exponential model such as

$$E(y_{it}|\mathbf{x}_{it}, c_i) = c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta})$$
$$c_i \geq 0$$

usually makes more sense than a linear model. Plus, we cannot use $\log(y_{it})$ if $P(y_{it} = 0) > 0$.

- General idea is to use models that are logically consistent with the nature of $y_{it}$.

- Not a bad idea to start with a linear model. For example, if $y_{it}$ is binary, we use an unobserved effects linear probability model estimated by fixed effects.
- In comparing across models it is important not to get tripped up by focusing on parameters. Estimating partial effects (magnitudes, not just directions) should be the focus in most applications.

## 2. CRE versus Other Approaches

- CRE contains traditional random effects as a special case. Can test the key RE assumption that heterogeneity is independent of time-varying covariates.
- Conditional MLE, which is used to eliminate unobserved heterogeneity, can be applied only in special cases. Even when it can, it usually relies on strong independence assumptions.
- "Fixed Effects," where the $c_i$ are treated as parameters to estimate, usually suffers from an incidental parameters problem. Recent work on adjustments for "large" $T$ seem promising but has drawbacks.

| | FE | CMLE | CRE |
|---|---|---|---|
| Restricts $D(c_i \| x_i)$? | No | No | Yes |
| Incidental Parameters with Small $T$? | Yes | No | No |
| Restricts Time Series Dependence/Heterogeneity? | Yes (1) | Yes (2) | No |
| Only Special Models? | No (3) | Yes | No |
| APEs Identified? | Yes (4) | No | Yes |
| Unbalanced Panels? | Yes | Yes | Yes (5) |
| Can Estimate $D(c_i)$? | Yes (4) | No | Yes (6) |

1. The large $T$ approximations, including bias adjustments, assume weak dependence and often stationarity.
2. Usually conditional independence, unless estimator is inherently fully robust (linear, Poisson).
3. Need at least one more time period than sources of heterogeneity.
4. Subject to the incidental parameters problem.
5. Subject to exchangeability restrictions.
6. Usually requires conditional independence or some other restriction.

## 3. Nonlinear Unobserved Effects Models

- Consider an unobserved effects probit model:

$$P(y_{it} = 1|\mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i),\ t = 1,\ldots,T,$$

where $\Phi(\bullet)$ is the standard normal cdf and $\mathbf{x}_{it}$ is $1 \times K$.

- Logit replaces $\Phi(z)$ with $\Lambda(z) = \exp(z)/[1 + \exp(z)]$.

• What are the quantities of interest? In economics, usually partial effects.

• For a continuous $x_{tj}$, the partial effect is

$$\frac{\partial P(y_t = 1|\mathbf{x}_t, c)}{\partial x_{tj}} = \beta_j \phi(\mathbf{x}_t \boldsymbol{\beta} + c),$$

where $\phi(\cdot)$ is the standard normal pdf.

• This partial effect (PE) depends on the values of the all observed covariates, and on the unobserved heterogeneity value $c$.

• The sign of the PE is the same as the sign of $\beta_j$, but we usually want the magnitude.

• If we have two continuous variables, the ratio of the partial effects is constant and equal to the ratio of coefficients:

$$\frac{\beta_j \phi(\mathbf{x}_t \boldsymbol{\beta} + c)}{\beta_h \phi(\mathbf{x}_t \boldsymbol{\beta} + c)} = \frac{\beta_j}{\beta_h}$$

• The ratio still does not tell us the size of the effect of each. And what about discrete covariates or more complicated functional forms (quadratics, interactions)?

• Discrete changes:

$$\Phi(\mathbf{x}_t^{(1)}\boldsymbol{\beta} + c) - \Phi(\mathbf{x}_t^{(0)}\boldsymbol{\beta} + c),$$

where $\mathbf{x}_t^{(0)}$ and $\mathbf{x}_t^{(1)}$ are set at different values. Again, this partial effect depends on $c$ (as well as the values of the covariates).

• Assuming we can consistently estimate $\boldsymbol{\beta}$, what should we do about the unobservable $c$?

- General Setup: Suppose we are interested in

$$E(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i) = m_t(\mathbf{x}_{it}, \mathbf{c}_i),$$

where $\mathbf{c}_i$ can be a vector of unobserved heterogeneity.

- Partial effects: If $x_{tj}$ is continuous, then its PE is

$$\theta_j(\mathbf{x}_t, \mathbf{c}) \equiv \frac{\partial m_t(\mathbf{x}_t, \mathbf{c})}{\partial x_{tj}}.$$

- Issues for discrete changes are similar.

• How do we account for unobserved $\mathbf{c}_i$? If we know enough about the distribution of $\mathbf{c}_i$ we can insert meaningful values for $\mathbf{c}$. For example, if $\boldsymbol{\mu}_{\mathbf{c}} = E(\mathbf{c}_i)$, then we can compute the *partial effect at the average (PEA)*,

$$PEA_j(\mathbf{x}_t) = \theta_j(\mathbf{x}_t, \boldsymbol{\mu}_{\mathbf{c}}) = \frac{\partial m_t(\mathbf{x}_t, \boldsymbol{\mu}_{\mathbf{c}})}{\partial x_{tj}}$$

Of course, we need to estimate the function $m_t$ and $\boldsymbol{\mu}_{\mathbf{c}}$.

• If we can estimate other features of the distribution of $\mathbf{c}_i$ we can insert different quantiles, or a certain number of standard deviations from the mean.

- An alternative measure is the *average partial effect* (APE) (or *population average effect*), obtained by averaging across the distribution of $\mathbf{c}_i$:

$$APE(\mathbf{x}_t) = E_{\mathbf{c}_i}[\theta_j(\mathbf{x}_t, \mathbf{c}_i)].$$

- The APE is closely related to the notion of the *average structural function (ASF)* [Blundell and Powell (2003, *REStud*)]. The ASF is defined as a function of $\mathbf{x}_t$:

$$ASF(\mathbf{x}_t) = E_{\mathbf{c}_i}[m_t(\mathbf{x}_t, \mathbf{c}_i)].$$

- Passing the derivative (with respect to $x_{tj}$) through the expectation in the ASF gives an APE.

• If

$$E(y_{it}|\mathbf{x}_{it},c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i)$$
$$c_i \sim Normal(0,\sigma_c^2)$$

can show that

$$PEA_j(\mathbf{x}_t) = \beta_j\phi(\mathbf{x}_t\boldsymbol{\beta})$$
$$APE_j(\mathbf{x}_t) = \beta_{cj}\phi(\mathbf{x}_t\boldsymbol{\beta}_c)$$

where $\boldsymbol{\beta}_c = \boldsymbol{\beta}/(1+\sigma_c^2)^{1/2}$.

• We can have $PEA_j(\mathbf{x}_t) < APE_j(\mathbf{x}_t)$ or $PEA_j(\mathbf{x}_t) > APE_j(\mathbf{x}_t)$ and the direction of the inequality can change with $\mathbf{x}_t$.

• If $c_i$ is independent of $\mathbf{x}_{it}$ we cannot estimate $\boldsymbol{\beta}$ but we can estimate the scaled vector, $\boldsymbol{\beta}_c$.

• Somewhat counterintuitive, but generally the APE is identified more often than the PEA.

• Example reveals that the "problem" of attenuation bias is a red herring. If we can estimate $\boldsymbol{\beta}_c$ we can get the signs of the PEs and relative effects. In addition, we can obtain the average partial effects.

- Important: Definitions of partial effects do not depend on whether $\mathbf{x}_{it}$ is correlated with $\mathbf{c}_i$. $\mathbf{x}_{it}$ could include contemporaneously endogenous variables or even $y_{i,t-1}$.
- Whether we can estimate the PEs certainly does depend on what we assume about the relationship beween $\mathbf{c}_i$ and $\{\mathbf{x}_{it}\}$.
- Focus on APEs means very general analyses are available – even nonparametric analyses.

- To summarize a partial effect as a single value, we need to deal with the presence $\mathbf{x}_t$.
- We can evaluate $\mathbf{x}_t$ at the sample average (for each $t$, say, or across all $t$). Or, we can average the partial effects across all $i$. More later.
- Stata has three commands, `mfx`, `margeff`, and (most recently) `margins`. Latter allows PEA or APE calculations (usually).

**Heterogeneity Distributions**

• With the CRE appoach we can, under enough assumptions, identify and consistently estimate the parameters in a *conditional* distribution $D(\mathbf{c}_i|\mathbf{w}_i)$ for some observed vector $\mathbf{w}_i$.

• Let $f(\mathbf{c}|\mathbf{w};\boldsymbol{\gamma})$ denote the identified conditional density and let $g(\mathbf{c})$ be the unconditional density. Then

$$\hat{g}(\mathbf{c}) = N^{-1}\sum_{i=1}^{N} f(\mathbf{c}|\mathbf{w}_i;\hat{\boldsymbol{\gamma}})$$

is a consistent estimator of $g(\mathbf{c})$. See Wooldridge (2011, Economics Letters).

## 4. Assumptions

- The CRE approach typically relies on three kinds of assumptions:

1. How do idiosyncratic (time-varying) shocks (which may be serially correlated) relate to the history of covariates, $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$?

2. Conditional Independence (which effectively rules out serial correlation in underlying shocks) or some other specific form of dependence.

3. How does unobserved (time-constant) heterogeneity relate to $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$?

**Assumptions Relating $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$ and Shocks**

• As in linear case, we cannot get by with just specifying a model for the contemporaneous conditional distribution, $D(\mathbf{y}_{it} | \mathbf{x}_{it}, \mathbf{c}_i)$.

• For example, it is not nearly enough to just specify $P(y_{it} = 1 | \mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i)$.

• A general definition of strict exogeneity (conditional on the heterogeneity) models is

$$D(y_{it} | \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \mathbf{c}_i) = D(y_{it} | \mathbf{x}_{it}, \mathbf{c}_i).$$

• In some cases strict exogeneity in the conditional mean sense is sufficient.

- There is a sequential exogeneity assumption, too. Dynamic models come later.
- Neither strict nor sequential exogeneity allows for contemporaneous endogeneity of one or more elements of $\mathbf{x}_{it}$, where, say, $x_{itj}$ is correlated with unobserved, time-varying unobservables that affect $y_{it}$.

**Conditional Independence**

- In linear models, serial dependence of idiosyncratic shocks is easily dealt with, usually by “cluster robust” inference with RE or FE.
- Or, we can use a GLS method. In the linear case with strictly exogenous covariates, serial correlation never results in inconsistent estimation, even if improperly modeled.
- The situation is different with nonlinear models estimated by full MLE: If independence is used it is usually needed for consistency.

- Conditional independence (CI) (with strict exogeneity imposed):

$$D(y_{i1}, \ldots, y_{iT} | \mathbf{x}_i, \mathbf{c}_i) = \prod_{t=1}^{T} D(y_{it} | \mathbf{x}_{it}, \mathbf{c}_i).$$

- Even after conditioning on $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$ we observe serial correlation in $\{y_{it}\}$ due to $\mathbf{c}_i$; but only due to $\mathbf{c}_i$.

- CI rules out shocks affecting $y_{it}$ being serially correlated. For example, if we write a binary response as

$$y_{it} = 1[\mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} \geq 0],$$

the $\{u_{it}\}$ would have to be serially independent for CI to hold.

- Unlike linear estimation, joint MLEs that use the serial independence assumption in estimation are usually inconsistent when the assumption fails.
- Unless it has been shown otherwise, one should assume CI is needed for consistency.

- In the CRE framework, CI plays a critical role in being able to estimate the "structural" parameters and the parameters in the distribution of $\mathbf{c}_i$ (and, therefore, in estimating PEAs).
- In a broad class of popular models, CI plays no essential role in estimating APEs using pooled methods (and GLS-type variants).

**Assumptions Relating $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$ and Heterogeneity**

*Random Effects*

• Generally stated, the key RE assumption is

$$D(\mathbf{c}_i|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) = D(\mathbf{c}_i).$$

and then the unconditional distribution of $\mathbf{c}_i$ is modeled. This is very strong.

• An implication of independence between $\mathbf{c}_i$ and $\mathbf{x}_i$ is that all APEs can be obtained by just estimating $E(y_{it}|\mathbf{x}_{it} = \mathbf{x}_t)$, that is, by ignoring the heterogeneity entirely.

• In the unobserved effects probit model, if $c_i$ is independent of $\{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}\}$ with a $Normal(0, \sigma_c^2)$ distribution, it can be show that

$$P(y_{it} = 1|\mathbf{x}_{it}) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta}_c),$$

where $\boldsymbol{\beta}_c = \boldsymbol{\beta}/(1 + \sigma_c^2)^{1/2}$.

• As discussed earlier, these scaled coefficients are actually what we want because they index the APEs.

*Correlated Random Effects*

• A CRE framework allows dependence between $\mathbf{c}_i$ and $\mathbf{x}_i$, but it is restricted in some way.

• In a parametric setting, we specify a distribution for $D(\mathbf{c}_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})$, as in Chamberlain (1980,1982), and much work since.

• Distributional assumptions that lead to simple estimation – homoskedastic normal with a linear conditional mean — are, in principle, restrictive. (However, estimates of average partial effects can be pretty resilient.)

- A general nonparametric assumption is

$$D(\mathbf{c}_i|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) = D(\mathbf{c}_i|\bar{\mathbf{x}}_i),$$

which conserves on degrees of freedom and often makes sense. APEs are identified very generally under this restriction.

- Often $D(\mathbf{c}_i|\mathbf{x}_i) = D(\mathbf{c}_i|\bar{\mathbf{x}}_i)$ is used in conjunction with flexible parametric models for $D(\mathbf{c}_i|\bar{\mathbf{x}}_i)$.

• We can directly include explanatory variables that do not change over time (but we may not be able to estimate their "causal" effects).

• Especially with larger $T$ the CRE approach can be flexible. We can allow $D(\mathbf{c}_i|\mathbf{x}_i)$ to depend on individual-specific trends or measures of dispersion in $\{\mathbf{x}_{it} : t = 1,\ldots,T\}$.

*Fixed Effects*

• The label "fixed effects" is used in different ways.

1. The $\mathbf{c}_i,\ i = 1,\ldots,N$ are parameters to be estimated along with fixed parameters. Usually leads to an "incidental parameters problem" unless $T$ is "large."

• Recent work on bias adjustments for both parameters and APEs. But time series dependence and heterogeneity are restricted.

2. $D(\mathbf{c}_i|\mathbf{x}_i)$ is unrestricted and we look for objective functions that do not depend on $\mathbf{c}_i$ but still identify the population parameters. Leads to "conditional MLE" (CMLE) if we can find sufficient statistics $\mathbf{s}_i$ such that

$$D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i,\mathbf{c}_i,\mathbf{s}_i) = D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i,\mathbf{s}_i)$$

where this latter distribution still depends on the constant parameters.

- In the rare case where CMLE is applicable, conditional independence is usually maintained – in particular, for unobserved effects logit models.
- Essentially by construction, PEAs and APEs are generally unidentified by methods that use conditioning to eliminate $\mathbf{c}_i$. We can get directions and (sometimes) relative magnitudes, or effects on log-odds, but not average partial effects.

## 5. Correlated Random Effects Probit

- The model is

$$P(y_{it} = 1|\mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i),\ t = 1, \ldots, T.$$

- Strict exogeneity conditional on $c_i$:

$$P(y_{it} = 1|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, c_i) = P(y_{it} = 1|\mathbf{x}_{it}, c_i),\ t = 1, \ldots, T.$$

- Conditional independence [where we condition on $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$ and $c_i$]:

$$D(y_{i1}, \ldots, y_{iT}|\mathbf{x}_i, c_i) = D(y_{i1}|\mathbf{x}_i, c_i) \cdots D(y_{iT}|\mathbf{x}_i, c_i)$$

- Model for $D(c_i|\mathbf{x}_i)$:

$$c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i,\ a_i|\mathbf{x}_i \sim Normal(0,\sigma_a^2).$$

- Chamberlain: Replace $\bar{\mathbf{x}}_i$ with $\mathbf{x}_i = (\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})$.

- Can obtain the first three assumptions from a latent variable model:

$$\begin{aligned} y_{it} &= 1[\mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} > 0] \\ u_{it}|(\mathbf{x}_{it},c_i) &\sim Normal(0,1) \\ D(u_{it}|\mathbf{x}_i,c_i) &= D(u_{it}|\mathbf{x}_{it},c_i) \\ \{u_{it} &: t = 1,\ldots,T\} \text{ independent} \end{aligned}$$

- Can include time dummies in $\mathbf{x}_{it}$ (but omit from $\bar{\mathbf{x}}_i$). Can also include time-constant elements as extra controls.

- If $\boldsymbol{\xi} = \mathbf{0}$, get the traditional random effects probit model.

- MLE (conditional on $\mathbf{x}_i$) is relatively straightforward. Under the assumption of *iid* normal shocks it is based on the joint distribution $D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i)$.

- In Stata:

```
egen x1bar = mean(x1), by(id)
⋮
egen xKbar = mean(xK), by(id)
xtprobit y x1 ... xK x1bar ... xKbar d2 ... dT,
re
```

• With conditional independence we can estimate features of the unconditional distribution of $c_i$.

• For example,

$$\hat{\mu}_c = \hat{\psi} + \bar{\mathbf{x}}\hat{\boldsymbol{\xi}}$$

$$\hat{\sigma}_c^2 \equiv \hat{\boldsymbol{\xi}}'\left(N^{-1}\sum_{i=1}^{N}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})\right)\hat{\boldsymbol{\xi}} + \hat{\sigma}_a^2$$

• Can evaluate PEs at, say, the estimated mean value, say $\hat{\mu}_c$, or look at $\hat{\mu}_c \pm k\hat{\sigma}_c$ for various $k$. Can plug in mean values of $\mathbf{x}_t$, too, or other specific values.

• As shown in Wooldridge (2011, Economics Letters), the unconditional heterogeneity distribution is consistently estimated as

$$\hat{g}(c) = N^{-1} \sum_{i=1}^{N} \phi[(c - \hat{\psi} - \bar{\mathbf{x}}_i \hat{\boldsymbol{\xi}})/\hat{\sigma}_a]/\hat{\sigma}_a,\ c \in \mathbb{R}$$

• The APEs are identified from from the average structural function, easily estimated as

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1}\sum_{i=1}^{N}\Phi(\mathbf{x}_t\hat{\boldsymbol{\beta}}_a + \hat{\psi}_a + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}}_a)$$

• The scaled coefficients are, for example, $\hat{\boldsymbol{\beta}}_a = \hat{\boldsymbol{\beta}}/(1+\hat{\sigma}_a^2)^{1/2}$.

• Take derivatives and changes with respect to $\mathbf{x}_t$. Can further average out across $\mathbf{x}_{it}$ to get a single APE.

• In Stata, `margeff` evaluates the heterogeneity at the mean (when the heterogeneity is independent of the covariates) but then averages the partial effects across the covariates.

• Conditional independence is strong, and the usual RE estimator not known to be robust to its violation. (Contrast RE estimation of the linear model.)

• If we focus on APEs, can just use a pooled probit method and completely drop the serial independence assumption.

• Pooled probit estimates the scaled coefficients directly because

$$P(y_{it} = 1|\mathbf{x}_i) = P(y_{it} = 1|\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta}_a + \psi_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a).$$

- In Stata, pooled probit and obtaining marginal effects are straightforward:

```
egen x1bar = mean(x1), by(id)
⋮
egen xKbar = mean(xK), by(id)
probit y x1 ... xK x1bar ... xKbar d2 ... dT,
cluster(id)
margeff
margins, dydx(*)
```

- Pooled probit is inefficient relative to CRE probit.
- We can try to get back some of the efficiency loss by using "generalized estimating equations" (GEE), which is essentially multivariate nonlinear least squares.

```
xtgee y x1 ... xK x1bar ... xKbar d2 ... dT,
fam(bin) link(probit) corr(exch) robust
```

- GEE might be more efficient than pooled probit, but there is no guarantee. It is as robust as pooled probit.
- GEE is less efficient than full MLE under serial independence, but the latter is less robust.

- As shown in Papke and Wooldridge (2008, Journal of Econometrics), if $y_{it}$ is a fraction we can use either pooled probit or GEE (but not full MLE) without any change to the estimation.
- With $0 \leq y_{it} \leq 1$ we start with

$$E(y_{it}|\mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i).$$

- When the heterogeneity is integrated out,

$$E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta}_a + \psi_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a).$$

- Now exploit that the Bernoulli distribution is in the linear exponential family. Pooled “probit” is a now a pooled quasi-MLE. Make inference fully robust, as before. Marginal effects calculations are unchanged.
- Can also use GEE with the probit response function as the mean but in a feasible GLS estimation, where the conditional variance-covariance matrix has constant correlations and is clearly misspecified.

```
glm y x1 ... xK x1bar ... xKbar d2 ... dT,
fam(bin) link(probit) cluster(id)
margins, dydx(*)
xtgee y x1 ... xK x1bar ... xKbar d2 ... dT,
fam(bin) link(probit) corr(exch) cluster(id)
```

**EXAMPLE**: Married Women's Labor Force Participation, LFP.DTA

```
. des lfp kids hinc

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------------------------
lfp             byte   %9.0g                  =1 if in labor force
kids            byte   %9.0g                  number children < 18
hinc            float  %9.0g                  husband's monthly income, $

. tab period

  1 through |
  5, each 4 |
months long |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      5,663       20.00       20.00
          2 |      5,663       20.00       40.00
          3 |      5,663       20.00       60.00
          4 |      5,663       20.00       80.00
          5 |      5,663       20.00      100.00
------------+-----------------------------------
      Total |     28,315      100.00

. egen kidsbar = mean(kids), by(id)

. egen lhincbar = mean(lhinc), by(id)
```

| LFP | (1) | (2) | | (3) | | (4) | | (5) |
|---|---|---|---|---|---|---|---|---|
| Model | Linear | Probit | | CRE Probit | | CRE Probit | | FE Logit |
| Est. Method | FE | Pooled MLE | | Pooled MLE | | MLE | | MLE |
| | Coef. | Coef. | APE | Coef. | APE | Coef. | APE | Coef. |
| *kids* | −.0389 | −.199 | −.0660 | −.117 | −.0389 | −.317 | −.0403 | −.644 |
| | (.0092) | (.015) | (.0048) | (.027) | (.0085) | (.062) | (.0104) | (.125) |
| *lhinc* | −.0089 | −.211 | −.0701 | −.029 | −.0095 | −.078 | −.0099 | −.184 |
| | (.0046) | (.024) | (.0079) | (.014) | (.0048) | (.041) | (.0055) | (.083) |
| $\overline{kids}$ | — | — | — | −.086 | — | −.210 | — | — |
| | — | — | — | (.031) | — | (.071) | — | — |
| $\overline{lhinc}$ | — | — | — | −.250 | — | −.646 | — | — |
| | — | — | — | (.035) | — | (.079) | — | — |

```
. * Linear model by FE:

. xtreg lfp kids lhinc per2-per5, fe cluster(id)

Fixed-effects (within) regression               Number of obs      =     28315
Group variable (i): id                          Number of groups   =      5663


                                 (Std. Err. adjusted for 5663 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         lfp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.0388976   .0091682    -4.24   0.000    -.0568708   -.0209244
       lhinc |  -.0089439   .0045947    -1.95   0.052    -.0179513    .0000635
        per2 |  -.0042799    .003401    -1.26   0.208    -.0109472    .0023875
        per3 |  -.0108953   .0041859    -2.60   0.009    -.0191012   -.0026894
        per4 |  -.0123002   .0044918    -2.74   0.006    -.0211058   -.0034945
        per5 |  -.0176797   .0048541    -3.64   0.000    -.0271957   -.0081637
       _cons |   .8090216   .0375234    21.56   0.000     .7354614    .8825818
-------------+----------------------------------------------------------------
     sigma_u |  .42247488
     sigma_e |  .21363541
         rho |  .79636335   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. * Fixed Effects Logit:

. xtlogit lfp kids lhinc per2-per5, fe

Conditional fixed-effects logistic regression   Number of obs      =      5275
Group variable: id                              Number of groups   =      1055

                                                Obs per group: min =         5
                                                               avg =       5.0
                                                               max =         5

                                                LR chi2(6)         =     57.27
Log likelihood  = -2003.4184                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.6438386   .1247828    -5.16   0.000    -.8884084   -.3992688
       lhinc |  -.1842911   .0826019    -2.23   0.026    -.3461878   -.0223943
        per2 |  -.0928039   .0889937    -1.04   0.297    -.2672283    .0816205
        per3 |  -.2247989   .0887976    -2.53   0.011     -.398839   -.0507587
        per4 |  -.2479323   .0888953    -2.79   0.005     -.422164   -.0737006
        per5 |  -.3563745   .0888354    -4.01   0.000    -.5304886   -.1822604
------------------------------------------------------------------------------

. di 644/184
3.5

. di 389/89
4.3707865
```

```
. * CRE probit:

. xtprobit lfp kids lhinc kidsbar lhincbar educ black age agesq per2-per5, re

Random-effects probit regression                Number of obs      =     28315
Group variable (i): id                          Number of groups   =      5663

                                                Wald chi2(12)      =    824.11
Log likelihood  = -8990.0898                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.3174051     .06203    -5.12   0.000    -.4389816   -.1958287
       lhinc |  -.0777949   .0414033    -1.88   0.060    -.1589439    .0033541
     kidsbar |  -.2098409   .0708676    -2.96   0.003    -.3487389   -.0709429
    lhincbar |  -.6463674   .0792719    -8.15   0.000    -.8017374   -.4909974
        educ |    .221596   .0147891    14.98   0.000     .1926099    .2505821
       black |   .5226558   .1502331     3.48   0.001     .2282042    .8171073
         age |   .4036543   .0287538    14.04   0.000     .3472979    .4600107
       agesq |  -.0054898   .0003536   -15.52   0.000    -.0061829   -.0047966
        per2 |   -.034359   .0438562    -0.78   0.433    -.1203156    .0515976
        per3 |  -.0954482   .0439688    -2.17   0.030    -.1816253    -.009271
        per4 |  -.1046944   .0439108    -2.38   0.017    -.1907581   -.0186308
        per5 |  -.1559446   .0435241    -3.58   0.000    -.2412502   -.0706389
       _cons |  -2.080352   .6567295    -3.17   0.002    -3.367518   -.7931854
-------------+----------------------------------------------------------------
    /lnsig2u |    1.73677   .0266277                      1.684581     1.78896
-------------+----------------------------------------------------------------
     sigma_u |   2.383059   .0317277                      2.321679    2.446063
         rho |   .8502764   .0033899                      .8435102    .8567997
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  1.5e+04 Prob >= chibar2 = 0.000
```

```
. predict xdhat, xb

. gen xdhata = xdhat/sqrt(1 + 2.383059^2)

. di 1/sqrt(1 + 2.383059^2)
.38694144

. * Scaled coefficients to compare with pooled probit:

. di (1/sqrt(1 + 2.383059^2))*_b[kids]
-.1228172

. di (1/sqrt(1 + 2.383059^2))*_b[lhinc]
-.03010209
```

```
. probit lfp kids lhinc kidsbar lhincbar educ black age agesq per2-per5,
        cluster(id)

Probit regression                                 Number of obs   =      28315
                                                  Wald chi2(12)   =     538.09
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -16516.436                 Pseudo R2       =     0.0673

                                 (Std. Err. adjusted for 5663 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.1173749   .0269743    -4.35   0.000    -.1702435   -.0645064
       lhinc |  -.0288098    .014344    -2.01   0.045    -.0569234   -.0006961
     kidsbar |  -.0856913   .0311857    -2.75   0.006     -.146814   -.0245685
    lhincbar |  -.2501781   .0352907    -7.09   0.000    -.3193466   -.1810097
        educ |   .0841338   .0067302    12.50   0.000     .0709428    .0973248
       black |   .2030668   .0663945     3.06   0.002     .0729359    .3331976
         age |   .1516424   .0124831    12.15   0.000      .127176    .1761089
       agesq |  -.0020672   .0001553   -13.31   0.000    -.0023717   -.0017628
        per2 |  -.0135701   .0103752    -1.31   0.191    -.0339051    .0067648
        per3 |  -.0331991   .0127197    -2.61   0.009    -.0581293    -.008269
        per4 |  -.0390317   .0136244    -2.86   0.004    -.0657351   -.0123284
        per5 |  -.0552425   .0146067    -3.78   0.000    -.0838711   -.0266139
       _cons |  -.7260562   .2836985    -2.56   0.010    -1.282095   -.1700173
------------------------------------------------------------------------------
```

```
. drop xdhat xdhata

. predict xdhat, xb

. gen scale = normden(xdhat)

. sum scale

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       scale |     28315    .3310079     .057301   .0694435   .3989423

. di .331*(-.117375)
-.03885113

. di .331*(-.02881)
-.00953611
```

```
. margeff

Average marginal effects on Prob(lfp==1) after probit

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |   -.038852   .0089243    -4.35   0.000    -.0563433   -.0213608
       lhinc |  -.0095363   .0047482    -2.01   0.045    -.0188426     -.00023
     kidsbar |  -.0283645   .0102895    -2.76   0.006    -.0485315   -.0081974
    lhincbar |  -.0828109   .0115471    -7.17   0.000    -.1054428    -.060179
        educ |    .027849   .0021588    12.90   0.000     .0236178    .0320801
       black |   .0643443   .0200207     3.21   0.001     .0251043    .1035842
         age |   .0501948   .0039822    12.60   0.000     .0423898    .0579998
       agesq |  -.0006843   .0000493   -13.88   0.000    -.0007809   -.0005876
        per2 |  -.0044999   .0034482    -1.30   0.192    -.0112583    .0022585
        per3 |  -.0110375   .0042512    -2.60   0.009    -.0193698   -.0027052
        per4 |  -.0129865   .0045606    -2.85   0.004    -.0219252   -.0040479
        per5 |  -.0184197   .0049076    -3.75   0.000    -.0280385    -.008801
------------------------------------------------------------------------------
```

```
. probit lfp kids lhinc educ black age agesq per2-per5, cluster(id)

Probit regression                                 Number of obs   =      28315
                                                  Wald chi2(10)   =     537.36
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -16556.671                 Pseudo R2       =     0.0651

                                  (Std. Err. adjusted for 5663 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.1989144   .0153153   -12.99   0.000    -.2289319   -.1688969
       lhinc |  -.2110739   .0242901    -8.69   0.000    -.2586816   -.1634661
        educ |   .0796863   .0065453    12.17   0.000     .0668577    .0925149
       black |   .2209396   .0659041     3.35   0.001       .09177    .3501093
         age |   .1449159   .0122179    11.86   0.000     .1209693    .1688624
       agesq |  -.0019912   .0001522   -13.08   0.000    -.0022895   -.0016928
        per2 |  -.0124245   .0104551    -1.19   0.235    -.0329162    .0080672
        per3 |  -.0325178   .0127431    -2.55   0.011    -.0574938   -.0075418
        per4 |   -.046097   .0136286    -3.38   0.001    -.0728087   -.0193853
        per5 |  -.0577767    .014632    -3.95   0.000    -.0864548   -.0290985
       _cons |  -1.064449    .261872    -4.06   0.000    -1.577709   -.5511895
------------------------------------------------------------------------------
```

```
. margeff

Average marginal effects on Prob(lfp==1) after probit

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        kids |  -.0660184   .0049233   -13.41   0.000    -.0756678    -.056369
       lhinc |   -.070054   .0079819    -8.78   0.000    -.0856981   -.0544099
        educ |   .0264473   .0021119    12.52   0.000     .0223082    .0305865
       black |   .0698835   .0197251     3.54   0.000      .031223     .108544
         age |   .0480966   .0039216    12.26   0.000     .0404105    .0557828
       agesq |  -.0006609   .0000486   -13.60   0.000    -.0007561   -.0005656
        per2 |  -.0041304   .0034828    -1.19   0.236    -.0109565    .0026957
        per3 |   -.010839   .0042694    -2.54   0.011    -.0192069   -.0024712
        per4 |  -.0153921   .0045809    -3.36   0.001    -.0243705   -.0064137
        per5 |  -.0193224   .0049309    -3.92   0.000    -.0289867   -.0096581
------------------------------------------------------------------------------

. * So, without accounting for heterogeneity through the time averages,
. * the effects are much larger.
```

```
. do ex15_5_boot1

. version 9

. capture program drop probit_boot

. program probit_boot, rclass
  1.
. probit lfp kids lhinc kidsbar lhincbar educ black age agesq per2-per5,
        cluster(id)
  2.
. predict xdhat, xb
  3. gen scale=normden(xdhat)
  4. gen pe1=scale*_b[kids]
  5. summarize pe1
  6. return scalar ape1=r(mean)
  7. gen pe2=scale*_b[lhinc]
  8. summarize pe2
  9. return scalar ape2=r(mean)
 10.
.
. drop xdhat scale pe1 pe2
 11. end
.
. bootstrap r(ape1) r(ape2), reps(500) seed(123) cluster(id) idcluster
            (newid): probit_boot
(running probit_boot on estimation sample)

Bootstrap replications (500)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

..................................................   500

Bootstrap results                               Number of obs      =     28315
                                                Number of clusters =      5663
                                                Replications       =       500
```

```
      command:  probit_boot
        _bs_1:  r(ape1)
        _bs_2:  r(ape2)

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _bs_1 |   -.038852   .0085179    -4.56   0.000    -.0555469   -.0221572
       _bs_2 |  -.0095363     .00482    -1.98   0.048    -.0189833   -.0000893
------------------------------------------------------------------------------

. program drop probit_boot

end of do-file

. do ex15_5_boot2

. capture program drop probit_boot

. program probit_boot, rclass
  1.
. probit lfp kids lhinc educ black age agesq per2-per5, cluster(id)
  2.
. predict xdhat, xb
  3. gen scale=normden(xdhat)
  4. gen pe1=scale*_b[kids]
  5. summarize pe1
  6. return scalar ape1=r(mean)
  7. gen pe2=scale*_b[lhinc]
  8. summarize pe2
  9. return scalar ape2=r(mean)
 10.
.
. drop xdhat scale pe1 pe2
 11. end

. bootstrap r(ape1) r(ape2), reps(500) seed(123) cluster(id) idcluster(newid):
```

```
            probit_boot
(running probit_boot on estimation sample)

Bootstrap replications (500)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

..................................................   500

Bootstrap results                               Number of obs      =     28315
                                                Number of clusters =      5663
                                                Replications       =       500

      command:  probit_boot
        _bs_1:  r(ape1)
        _bs_2:  r(ape2)

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _bs_1 |  -.0660184   .0047824   -13.80   0.000    -.0753916   -.0566451
       _bs_2 |   -.070054   .0078839    -8.89   0.000    -.0855061   -.0546019
------------------------------------------------------------------------------

. program drop probit_boot

end of do-file
```

- Many useful embellishments. For example, we can allow

$$c_i|\mathbf{x}_i \sim Normal[\psi + \bar{\mathbf{x}}_i\boldsymbol{\xi}, \sigma_a^2 \exp(\bar{\mathbf{x}}_i\boldsymbol{\zeta})],$$

and then use a version of "heteroskedastic probit" (probably pooled, but could use full MLE under conditional independence).

- If use the pooled method, applies if $y_{it}$ is binary or fractional.

- Estimation of APEs is based on

$$E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = \Phi[(1 + \sigma_a^2 \exp(\bar{\mathbf{x}}_i\boldsymbol{\zeta}))^{-1/2}(\mathbf{x}_{it}\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi})]$$

still straightforward. For continuous $x_{tj}$,

$$\widehat{APE_j}(\mathbf{x}_t) = \hat{\beta}_j \left\{ N^{-1} \sum_{i=1}^{N} (1 + \hat{\sigma}_a^2 \exp(\bar{\mathbf{x}}_i\hat{\boldsymbol{\zeta}}))^{-1/2} \cdot \phi[(1 + \hat{\sigma}_a^2 \exp(\bar{\mathbf{x}}_i\hat{\boldsymbol{\zeta}}))^{-1/2}(\mathbf{x}_t\hat{\boldsymbol{\beta}}_a + \hat{\psi}_a + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}}_a)] \right\}$$

- See Wooldridge (2010, Chapter 15).

## 6. CRE Tobit Model

- Unobserved effects Tobit model for a corner at zero is

$$y_{it} = \max(0, \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it})$$
$$D(u_{it}|\mathbf{x}_{it}, c_i) = Normal(0, \sigma_u^2)$$

- Strict exogeneity conditional on $c_i$:

$$D(u_{it}|\mathbf{x}_i, c_i) = D(u_{it}|\mathbf{x}_{it}, c_i)$$

- Conditional independence: The $\{u_{it} : t = 1, \ldots, T\}$ are independent.

- Model for $D(c_i|\mathbf{x}_i)$:

$$c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i,\ a_i|\mathbf{x}_i \sim Normal(0, \sigma_a^2).$$

- Joint MLE (conditional on $\mathbf{x}_i$) is relatively straightforward. It is based on the joint distribution $D(y_{i1}, \ldots, y_{iT}|\mathbf{x}_i)$.

- In Stata, called `xttobit` with the `re` option:

```
xttobit y x1 x2 ... xK x1bar ... xKbar, ll(0) re
```

- As in the probit case, we can estimate $\mu_c$ and $\sigma_c^2$:

$$\hat{\mu}_c = \hat{\psi} + \bar{\mathbf{x}}\hat{\boldsymbol{\xi}}$$

$$\hat{\sigma}_c^2 = \hat{\boldsymbol{\xi}}'\left(N^{-1}\sum_{i=1}^{N}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})\right)\hat{\boldsymbol{\xi}} + \hat{\sigma}_a^2$$

- Same estimate of heterogeneity distribution works, too.

• We can evaluate the partial effects of the Tobit function, $m(\mathbf{x}_t\hat{\boldsymbol{\beta}} + c, \hat{\sigma}_u^2)$ at different values of $c$, including $\hat{\mu}_c$ and $\hat{\mu}_c \pm k\hat{\sigma}_c$.

• Take derivatives or changes with respect to $\mathbf{x}_t$. For a continuous variable,

$$\hat{\beta}_j \Phi[(\mathbf{x}_t\hat{\boldsymbol{\beta}} + c)/\hat{\sigma}_u]$$

• APEs can be estimated from the mean function for the Tobit:

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1}\sum_{i=1}^{n} m(\mathbf{x}_t\hat{\boldsymbol{\beta}} + \hat{\psi} + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}}, \hat{\sigma}_a^2 + \hat{\sigma}_u^2)$$

where $m(z,\sigma^2)$ is the mean function for a Tobit.

• Take derivatives and differences with respect to elements of $\mathbf{x}_t$. Panel bootstrap for inference.

• For a continuous $x_{tj}$,

$$\widehat{APE_j}(\mathbf{x}_t) = \hat{\beta}_j \left[ N^{-1} \sum_{i=1}^{N} \Phi[(\mathbf{x}_t\hat{\boldsymbol{\beta}} + \hat{\psi} + \bar{\mathbf{x}}_i\hat{\boldsymbol{\xi}})/(\hat{\sigma}_a^2 + \hat{\sigma}_u^2)^{1/2}] \right]$$

• To estimate the APEs it suffices to estimate the variance of the composite error, $\sigma_v^2 = \sigma_a^2 + \sigma_u^2$.

• If we drop the conditional independence assumption and allow and serial dependence in $\{u_{it}\}$ then we only have the marginal distributions

$$D(y_{it}|\mathbf{x}_i) = D(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = Tobit(\mathbf{x}_{it}\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi}, \sigma_v^2)$$

• So, we can apply pooled Tobit, ignoring the serial correlation, to estimate $\boldsymbol{\beta}, \psi, \boldsymbol{\xi}$, and $\sigma_v^2$. • • We use the previous formula for the APEs. We cannot estimate PEAs because $E(c_i)$ is not identified; neither is $\boldsymbol{\beta}$ nor $\sigma_u^2$.

```
. use psid80_92

. des hours  nwifeinc exper ch0_2 ch3_5 ch6_17

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------------------------
hours           int    %9.0g                  annual work hours
nwifeinc        float  %9.0g                  (faminc - wife’s labor income)/1000
exper           float  %9.0g                  years of workforce experience
ch0_2           byte   %9.0g                  number of children in FU, 0-2
ch3_5           byte   %9.0g                  number of children in FU, 3-5
ch6_17          byte   %9.0g                  number of children in FU, 6-17

. sum hours  nwifeinc exper ch0_2 ch3_5 ch6_17

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       hours |     11674    1130.995    888.2304          0       5168
    nwifeinc |     11674    34.22192    40.00195  -14.99792     1412.2
       exper |     11674    11.80465    7.743591          0    45.1315
       ch0_2 |     11674    .1351722    .3700769          0          3
       ch3_5 |     11674    .1776598      .42228          0          3
-------------+--------------------------------------------------------
      ch6_17 |     11674    .8222546    1.008326          0          6
```

```
. tab year

   80 to 92 |      Freq.     Percent        Cum.
------------+-----------------------------------
         80 |        898        7.69        7.69
         81 |        898        7.69       15.38
         82 |        898        7.69       23.08
         83 |        898        7.69       30.77
         84 |        898        7.69       38.46
         85 |        898        7.69       46.15
         86 |        898        7.69       53.85
         87 |        898        7.69       61.54
         88 |        898        7.69       69.23
         89 |        898        7.69       76.92
         90 |        898        7.69       84.62
         91 |        898        7.69       92.31
         92 |        898        7.69      100.00
------------+-----------------------------------
      Total |     11,674      100.00

. * First, linear FE:

. xtreg hours nwifeinc ch0_2 ch3_5 ch6_17 marr y81-y92, fe cluster(id)

Fixed-effects (within) regression               Number of obs      =     11674
Group variable (i): id                          Number of groups   =       898

R-sq:  within  = 0.0719                         Obs per group: min =        13
       between = 0.0936                                        avg =      13.0
       overall = 0.0855                                        max =        13

                                                F(17,11657)        =     15.72
corr(u_i, Xb)  = -0.0945                        Prob > F           =    0.0000

                                   (Std. Err. adjusted for 898 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
    nwifeinc |  -.7752375   .3429502    -2.26   0.024    -1.448316   -.1021593
       ch0_2 |  -342.3774   26.64763   -12.85   0.000    -394.6763   -290.0784
       ch3_5 |  -254.1283   25.87788    -9.82   0.000    -304.9165     -203.34
      ch6_17 |  -42.95787   14.88673    -2.89   0.004    -72.17475   -13.74099
        marr |  -634.8048   286.1714    -2.22   0.027    -1196.448    -73.1613
         y81 |  -4.819715   16.29731    -0.30   0.767    -36.80502    27.16559
         y82 |  -14.88765    21.1851    -0.70   0.482     -56.4658    26.69049
         y83 |   6.612531   22.49192     0.29   0.769    -37.53039    50.75545
         y84 |   93.79139   25.58646     3.67   0.000      43.5751    144.0077
         y85 |   88.73714   25.97019     3.42   0.001     37.76773    139.7065
         y86 |   82.66214   27.36886     3.02   0.003     28.94769    136.3766
         y87 |   64.28464   27.83649     2.31   0.021     9.652411    118.9169
         y88 |   63.79163   29.35211     2.17   0.030     6.184826    121.3984
         y89 |   72.98518   30.60838     2.38   0.017     12.91279    133.0576
         y90 |   71.24956   31.55331     2.26   0.024     9.322657    133.1765
         y91 |   64.67996   32.47097     1.99   0.047     .9520418    128.4079
         y92 |   16.01242   33.21255     0.48   0.630    -49.17093    81.19577
       _cons |    1786.02    247.297     7.22   0.000     1300.672    2271.368
-------------+----------------------------------------------------------------
     sigma_u |  701.66249
     sigma_e |  503.92334
         rho |  .65972225   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. * Compute time averages:

. egen nwifeincb = mean(nwifeinc), by(id)

. egen ch0_2b = mean(ch0_2), by(id)

. egen ch3_5b = mean(ch3_5), by(id)

. egen ch6_17b = mean(ch6_17), by(id)

. egen marrb = mean(marr), by(id)

. * Correlated RE Tobit:
```

```
. xttobit hours nwifeinc ch0_2 ch3_5 ch6_17 marr y81-y92 nwifeincb-marrb,
          ll(0)

Random-effects tobit regression                 Number of obs      =     11674
Group variable (i): id                          Number of groups   =       898

Random effects u_i ~Gaussian                   Obs per group: min =        13
                                                               avg =      13.0
                                                               max =        13

                                                Wald chi2(22)      =   1501.20
Log likelihood  = -70733.195                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -1.554228   .3816927    -4.07   0.000    -2.302332   -.8061243
       ch0_2 |   -472.088   23.03087   -20.50   0.000    -517.2277   -426.9483
       ch3_5 |  -329.3896   19.49411   -16.90   0.000    -367.5974   -291.1819
      ch6_17 |  -46.11619   10.89609    -4.23   0.000    -67.47213   -24.76024
        marr |  -784.1809   155.0133    -5.06   0.000    -1088.001   -480.3604
         y81 |  -7.060588   31.52257    -0.22   0.823    -68.84369    54.72251
         y82 |   -38.9034   31.70009    -1.23   0.220    -101.0344    23.22764
         y83 |  -9.719573   31.68694    -0.31   0.759    -71.82483    52.38569
         y84 |   99.77618   31.61932     3.16   0.002     37.80345    161.7489
         y85 |   89.15912    31.7439     2.81   0.005     26.94222     151.376
         y86 |   82.60212   31.76385     2.60   0.009     20.34612    144.8581
         y87 |   48.59097   31.98439     1.52   0.129    -14.09729    111.2792
         y88 |   53.52189   32.09804     1.67   0.095    -9.389108    116.4329
         y89 |   68.69013   32.23667     2.13   0.033     5.507414    131.8728
         y90 |    71.2654    32.3657     2.20   0.028       7.8298     134.701
         y91 |   64.89096   32.48217     2.00   0.046     1.227067    128.5548
         y92 |   4.334129   32.82961     0.13   0.895    -60.01072    68.67898
   nwifeincb |  -7.639696   .6815067   -11.21   0.000    -8.975424   -6.303967
      ch0_2b |  -143.4709   155.0915    -0.93   0.355    -447.4448    160.5029
      ch3_5b |   531.2027    150.388     3.53   0.000     236.4475    825.9578
     ch6_17b |   5.854889   28.04159     0.21   0.835    -49.10563     60.8154
```

```
       marrb |   422.1631    161.491     2.61   0.009     105.6465    738.6796
       _cons |   1646.362   45.26091    36.37   0.000     1557.652    1735.072
-------------+----------------------------------------------------------------
    /sigma_u |   756.4032   10.45016    72.38   0.000     735.9213    776.8851
    /sigma_e |   621.7044    5.02536   123.71   0.000     611.8549    631.5539
-------------+----------------------------------------------------------------
         rho |   .5968169   .0069011                      .5832357    .6102823
------------------------------------------------------------------------------

  Observation summary:      3071  left-censored observations
                            8603     uncensored observations
                               0 right-censored observations

. testparm nwifeincb-marrb

 ( 1)  [hours]nwifeincb = 0
 ( 2)  [hours]ch0_2b = 0
 ( 3)  [hours]ch3_5b = 0
 ( 4)  [hours]ch6_17b = 0
 ( 5)  [hours]marrb = 0

           chi2(  5) =  165.08
         Prob > chi2 =    0.0000


. gen xbhata = xbhat/sqrt(756.4032^2 +  621.7044^2)

. gen PHIhata = norm(xbhata)

. sum PHIhata if y92

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     PHIhata |       898    .8367103    .0953704   .0029178   .9654008

-------------------------------------------------------------------------------------------------

. di  (.837)*(-1.554)
```

```
-1.300698

. di  (.837)*(-472.09)
-395.13933

. * Pooled Tobit with Time Averages:

. tobit hours nwifeinc ch0_2 ch3_5 ch6_17 marr y81-y92 nwifeincb-marrb, ll(0)

Tobit regression                                  Number of obs   =      11674
                                                  LR chi2(22)     =    1352.20
                                                  Prob > chi2     =     0.0000
Log likelihood = -75313.315                       Pseudo R2       =     0.0089

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -1.796524   .6073205    -2.96   0.003    -2.986975   -.6060744
       ch0_2 |  -491.6069   38.36112   -12.82   0.000    -566.8011   -416.4127
       ch3_5 |  -347.5099    32.7817   -10.60   0.000    -411.7675   -283.2523
      ch6_17 |  -48.12398   18.14746    -2.65   0.008    -83.69604   -12.55191
        marr |  -788.6605   257.0461    -3.07   0.002    -1292.514   -284.8071
         y81 |  -1.723103   52.74963    -0.03   0.974    -105.1212     101.675
         y82 |  -29.93459   52.90393    -0.57   0.572    -133.6352    73.76597
         y83 |   .1544965   52.88423     0.00   0.998    -103.5075    103.8165
         y84 |   111.7593   52.84133     2.11   0.034     8.181439    215.3372
         y85 |    98.8203   53.02693     1.86   0.062    -5.121366     202.762
         y86 |   91.11779   53.07409     1.72   0.086    -12.91632    195.1519
         y87 |   56.20641   53.35906     1.05   0.292    -48.38629    160.7991
         y88 |   58.45143   53.59859     1.09   0.275    -46.61078    163.5136
         y89 |   74.11085   53.83913     1.38   0.169    -31.42287    179.6446
         y90 |   77.83721   54.05111     1.44   0.150    -28.11203    183.7865
         y91 |   70.43439   54.27841     1.30   0.194    -35.96039    176.8292
         y92 |   4.969863   54.81622     0.09   0.928    -102.4791    112.4188
   nwifeincb |  -7.248981   .7293248    -9.94   0.000    -8.678579   -5.819382
      ch0_2b |   152.0109   124.2391     1.22   0.221    -91.51857    395.5403
      ch3_5b |   151.7502   118.9341     1.28   0.202    -81.38056     384.881
     ch6_17b |   44.11858   25.07548     1.76   0.079    -5.033552    93.27072
```

```
       marrb |   471.4367   259.4683     1.82   0.069    -37.16466    980.0381
       _cons |   1581.923   46.08447    34.33   0.000      1491.59    1672.256
-------------+----------------------------------------------------------------
      /sigma |   1079.331   8.836301                       1062.01    1096.651
------------------------------------------------------------------------------
  Obs. summary:       3071  left-censored observations at hours<=0
                      8603     uncensored observations
                         0 right-censored observations

. * These differ somewhat, but not in major ways, from the full MLEs.

. * Now drop the time averages, so RE Tobit:

. xttobit hours nwifeinc ch0_2 ch3_5 ch6_17 marr y81-y92, ll(0)

Random-effects tobit regression                 Number of obs      =     11674
Group variable (i): id                          Number of groups   =       898

Random effects u_i ~Gaussian                   Obs per group: min =        13
                                                              avg =      13.0
                                                              max =        13

                                                Wald chi2(17)      =   1222.37
Log likelihood  = -70782.086                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |   -2.25119   .3248083    -6.93   0.000    -2.887803   -1.614578
       ch0_2 |   -459.927   22.67389   -20.28   0.000    -504.3671    -415.487
       ch3_5 |  -313.4996   18.81897   -16.66   0.000    -350.3841   -276.6151
      ch6_17 |  -32.33052   9.819359    -3.29   0.001    -51.57611   -13.08493
        marr |  -657.5755   48.93306   -13.44   0.000    -753.4825   -561.6684
         y81 |  -6.015057   31.64666    -0.19   0.849    -68.04136    56.01125
         y82 |  -37.89952   31.82432    -1.19   0.234     -100.274    24.47499
         y83 |    -7.2714   31.78778    -0.23   0.819     -69.5743     55.0315
         y84 |   104.3436   31.71544     3.29   0.001     42.18249    166.5047
         y85 |   94.90622   31.82266     2.98   0.003     32.53496    157.2775
```

```
         y86 |   89.38999   31.84555     2.81   0.005     26.97386    151.8061
         y87 |    57.1533   32.03317     1.78   0.074    -5.630564    119.9372
         y88 |   64.08813   32.11484     2.00   0.046     1.144192    127.0321
         y89 |   81.55682   32.20542     2.53   0.011     18.43536    144.6783
         y90 |   85.75216   32.26838     2.66   0.008     22.50728     148.997
         y91 |   80.93763   32.36379     2.50   0.012     17.50576    144.3695
         y92 |   22.68549   32.63686     0.70   0.487    -41.28158    86.65255
       _cons |   1676.368   39.27514    42.68   0.000      1599.39    1753.346
-------------+----------------------------------------------------------------
    /sigma_u |   768.5483   12.40411    61.96   0.000     744.2367    792.8599
    /sigma_e |    624.285   5.068197   123.18   0.000     614.3515    634.2185
-------------+----------------------------------------------------------------
         rho |   .6024761   .0077085                      .5872944    .6175041
------------------------------------------------------------------------------

  Observation summary:      3071  left-censored observations
                            8603     uncensored observations
                               0 right-censored observations

. predict xbhat, xb

. gen xbhata = xbhat/sqrt(768.5483^2 +  624.285^2)

. gen PHIhata = normal(xbhata)

. sum PHIhata if y92

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     PHIhata |       898    .8240658    .0724009   .3761031   .9578886

. * The scale factor is similar, but the coefficient estimates are
. * somewhat different.
```

## 7. CRE Count Models

- The most common model for the conditional mean allows multiplicative in the heterogeneity:

$$E(y_{it}|\mathbf{x}_{it}, c_i) = c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta})$$

where $c_i \geq 0$ is the unobserved effect and $\mathbf{x}_{it}$ would incude a full set of year dummies in most cases.

- $y_{it}$ need not be a count for this mean to make sense. Should have $y_{it} \geq 0$.

- As in the linear case, standard estimation methods assume strict exogeneity of the covariates conditional on $c_i$:

$$E(y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i) = E(y_{it}|\mathbf{x}_{it},c_i).$$

- A very convenient and fully robust estimator is the Poisson conditional MLE (often called the “Poisson fixed effects” estimator).

- An alternative is to apply pooled Poisson, GEE Poisson, or Poisson RE approaches in a CRE setting. Model $c_i$ as

$$c_i = \exp(\psi + \bar{\mathbf{x}}_i \boldsymbol{\xi}) a_i$$

where $a_i$ is independent of $\mathbf{x}_i$ with unit mean. Then

$$E(y_{it}|\mathbf{x}_i) = \exp(\psi + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i \boldsymbol{\xi}).$$

- So, can use any common method and simply add $\bar{\mathbf{x}}_i$ as a set of covariates. Can add time-constant covariates, too.
- Can easily test $H_0 : \boldsymbol{\xi} = \mathbf{0}$.

- Stata commands:

```
glm y x1 ... xK x1bar ... xKbar, fam(poisson)
cluster(id)
xtgee y x1 ... xK x1bar ... xKbar, fam(poisson)
corr(uns) robust
xtpoisson y x1 ... xK x1bar ... xKbar, re
```

- Pooled Poisson and GEE only use $E(y_{it}|\mathbf{x}_i) = \exp(\psi + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi})$.

The Poisson RE method requires that $D(y_{it}|\mathbf{x}_i, c_i) \sim Poisson$, $a_i \sim Gamma(\delta, \delta)$, and conditional independence over time.

## 8. Nonparametric and Flexible Parametric Approaches

- Suppose strict exogeneity holds conditional on $\mathbf{c}_i$:

$$E(y_{it}|\mathbf{x}_i, \mathbf{c}_i) = E(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i) = m_t(\mathbf{x}_{it}, \mathbf{c}_i)$$

- But we do not want to use a parametric model for $D(\mathbf{c}_i|\mathbf{x}_i)$. Maybe we want to leave $m_t(\bullet, \bullet)$ unspecified, too.
- Altonji and Matzkin (2005, Econometrica) show how to identify the average structural function (and a local version) by using "exchangeability" assumptions on $D(\mathbf{c}_i|\mathbf{x}_i)$.

- Leading exchangeability assumption:

$$D(\mathbf{c}_i|\mathbf{x}_i) = D(\mathbf{c}_i|\bar{\mathbf{x}}_i)$$

- But $D(\mathbf{c}_i|\mathbf{x}_i)$ might depend also on the sample variance-covariance matrix,

$$\mathbf{S}_i = (T-1)^{-1}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$$

- Generally, let $\mathbf{w}_i$ be a set of exchangeable functions of $\{\mathbf{x}_{it}\}$ such that

$$D(\mathbf{c}_i|\mathbf{x}_i) = D(\mathbf{c}_i|\mathbf{w}_i)$$

- Under restrictions on the nature of $\mathbf{w}_i$, the ASF can be identified from

$$E(y_{it}|\mathbf{x}_i, \mathbf{w}_i) \equiv g_t(\mathbf{x}_{it}, \mathbf{w}_i).$$

$$ASF(\mathbf{x}_t) = E_{\mathbf{w}_i}[g_t(\mathbf{x}_t, \mathbf{w}_i)].$$

- Practically, might implement using flexible parametric forms for $g_t(\mathbf{x}_{it}, \mathbf{w}_i)$.

- For example, if $\mathbf{w}_i = \bar{\mathbf{x}}_i$ and $0 \leq y_{it} \leq 1$ we can just *start* with

$$E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = \Phi[\theta_t + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\gamma} + (\mathbf{x}_{it} \otimes \bar{\mathbf{x}}_i)\boldsymbol{\delta} + (\bar{\mathbf{x}}_i \otimes \bar{\mathbf{x}}_i)\boldsymbol{\eta}]$$

and estimate the parameters by pooled "probit" or a GLS-type procedure (GEE).

- For a continuous variable $x_{tj}$ the estimated APE is

$$N^{-1}\sum_{i=1}^{N}(\hat{\beta}_j + \bar{\mathbf{x}}_i\hat{\boldsymbol{\delta}}_j)\phi[\hat{\theta}_t + \mathbf{x}_t\hat{\boldsymbol{\beta}} + \bar{\mathbf{x}}_i\hat{\boldsymbol{\gamma}} + (\mathbf{x}_t \otimes \bar{\mathbf{x}}_i)\hat{\boldsymbol{\delta}} + (\bar{\mathbf{x}}_i \otimes \bar{\mathbf{x}}_i)\hat{\boldsymbol{\eta}}]$$

- If the model were linear, the pooled OLS estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ would be the FE estimates.

- For $y_{it} \geq 0$ can use

$$E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = \exp[\theta_t + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\gamma} + (\mathbf{x}_{it} \otimes \bar{\mathbf{x}}_i)\boldsymbol{\delta} + (\bar{\mathbf{x}}_i \otimes \bar{\mathbf{x}}_i)\boldsymbol{\eta}]$$

to allow for more heterogeneity than a single, multiplicative effect.

- In a parametric setting, we do not have to impose exchangeability in the CRE approach. For example, we can allow the unrestricted Chamberlain device or individual-specific trends in $\{\mathbf{x}_{it}\}$.
- Possibilities and quality of approximations have been barely explored. The nonparametric identification of APEs is liberating, because we can just start with flexible parametric models conditional on $(\mathbf{x}_{it}, \mathbf{w}_i)$ with $\mathbf{w}_i = \bar{\mathbf{x}}_i$ the leading but not only case.