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, $x_{it}$ is a set of observed explanatory variables, $c_i$ is heterogeneity. We are interested in the response probability as a function of $(x_t, c)$:

$$p(x_t, c) = P(y_{it} = 1 | x_{it} = x_t, c_i = c).$$

- Because $p(x_t, c)$ is a probability, a linear model, say

$$p(x_t, c) = x_t \beta + c,$$

can be a poor approximation.
• Or, suppose \( y_{it} \geq 0 \). An exponential model such as

\[
E(y_{it}|x_{it}, c_i) = c_i \exp(x_{it}\beta)
\]

\[
c_i \geq 0
\]

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.
<table>
<thead>
<tr>
<th></th>
<th>FE</th>
<th>CMLE</th>
<th>CRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricts $D(c_i</td>
<td>x_i)$?</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Incidental Parameters with Small $T$?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Restricts Time Series Dependence/Heterogeneity?</td>
<td>Yes$^{(1)}$</td>
<td>Yes$^{(2)}$</td>
<td>No</td>
</tr>
<tr>
<td>Only Special Models?</td>
<td>No$^{(3)}$</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>APEs Identified?</td>
<td>Yes$^{(4)}$</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Unbalanced Panels?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes$^{(5)}$</td>
</tr>
<tr>
<td>Can Estimate $D(c_i)$?</td>
<td>Yes$^{(4)}$</td>
<td>No</td>
<td>Yes$^{(6)}$</td>
</tr>
</tbody>
</table>

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:

\[ P(y_{it} = 1|x_{it}, c_i) = \Phi(x_{it}\beta + c_i), \ t = 1, \ldots, T, \]

where \( \Phi(\cdot) \) is the standard normal cdf and \( x_{it} \) 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 $x_{ij}$, the partial effect is

$$\frac{\partial P(y_t = 1 | x_t, c)}{\partial x_{ij}} = \beta_j \phi(x_i \beta + c),$$

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:

$$\frac{\beta_j \phi(x_t \beta + c)}{\beta_h \phi(x_t \beta + c)} = \frac{\beta_j}{\beta_h}$$

• 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:

\[ \Phi(x_t^{(1)} \beta + c) - \Phi(x_t^{(0)} \beta + c), \]

where \( x_t^{(0)} \) and \( x_t^{(1)} \) 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 \( \beta \), what should we do about the unobservable \( c \)?
• General Setup: Suppose we are interested in

\[ E(y_{it}|x_{it}, c_i) = m_t(x_{it}, c_i), \]

where \( c_i \) can be a vector of unobserved heterogeneity.

• Partial effects: If \( x_{tj} \) is continuous, then its PE is

\[ \theta_j(x_t, c) \equiv \frac{\partial m_t(x_t, c)}{\partial x_{tj}}. \]

• Issues for discrete changes are similar.
How do we account for unobserved $c_i$? If we know enough about the distribution of $c_i$ we can insert meaningful values for $c$. For example, if $\mu_c = E(c_i)$, then we can compute the partial effect at the average (PEA),

$$PEA_j(x_t) = \theta_j(x_t, \mu_c) = \frac{\partial m_t(x_t, \mu_c)}{\partial x_{tj}}$$

Of course, we need to estimate the function $m_t$ and $\mu_c$.

If we can estimate other features of the distribution of $c_i$ 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 $c_i$:

$$APE(x_t) = E_{c_i}[\theta_j(x_t, c_i)].$$

• The APE is closely related to the notion of the *average structural function* (ASF) [Blundell and Powell (2003, *RES*ud)]. The ASF is defined as a function of $x_t$:

$$ASF(x_t) = E_{c_i}[m_t(x_t, c_i)].$$

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

\[ E(y_{it}|x_{it}, c_i) = \Phi(x_{it}\beta + c_i) \]

\[ c_i \sim Normal(0, \sigma_c^2) \]

can show that

\[ PEA_j(x_t) = \beta_j\phi(x_t\beta) \]

\[ APE_j(x_t) = \beta_c j\phi(x_t\beta_c) \]

where \( \beta_c = \beta/(1 + \sigma_c^2)^{1/2} \).

• We can have \( PEA_j(x_t) < APE_j(x_t) \) or \( PEA_j(x_t) > APE_j(x_t) \) and the direction of the inequality can change with \( x_t \).
• If $c_i$ is independent of $x_{it}$ we cannot estimate $\beta$ but we can estimate the scaled vector, $\beta_c$.

• 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 $\beta_c$ 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 $x_{it}$ is correlated with $c_i$. $x_{it}$ 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 between $c_i$ and $\{x_{it}\}$.

• 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 $x_t$.

• We can evaluate $x_t$ 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 approach we can, under enough assumptions, identify and consistently estimate the parameters in a conditional distribution $D(c_i|w_i)$ for some observed vector $w_i$.

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

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

is a consistent estimator of $g(c)$. 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, \( \{x_{it} : t = 1, \ldots, T\} \)?
  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 \( \{x_{it} : t = 1, \ldots, T\} \)?
Assumptions Relating \( \{x_{it} : t = 1, \ldots, T\} \) and Shocks

- As in linear case, we cannot get by with just specifying a model for the contemporaneous conditional distribution, \( D(y_{it} | x_{it}, c_i) \).

- For example, it is not nearly enough to just specify

\[
P(y_{it} = 1 | x_{it}, c_i) = \Phi(x_{it} \beta + c_i).
\]

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

\[
D(y_{it} | x_{i1}, \ldots, x_{iT}, c_i) = D(y_{it} | x_{it}, c_i).
\]

- 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 $x_{it}$, 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):

\[
D(y_{i1}, \ldots, y_{iT}|x_i, c_i) = \prod_{t=1}^{T} D(y_{it}|x_{it}, c_i).
\]

• Even after conditioning on \( \{x_{it} : t = 1, \ldots, T\} \) we observe serial correlation in \( \{y_{it}\} \) due to \( c_i \); but only due to \( c_i \).
• CI rules out shocks affecting $y_{it}$ being serially correlated. For example, if we write a binary response as

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

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 $c_i$ (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 $\{x_{it} : t = 1,\ldots,T\}$ and Heterogeneity

Random Effects

• Generally stated, the key RE assumption is

$$D(c_i|x_{i1},\ldots,x_{iT}) = D(c_i).$$

and then the unconditional distribution of $c_i$ is modeled. This is very strong.

• An implication of independence between $c_i$ and $x_i$ is that all APEs can be obtained by just estimating $E(y_{it}|x_{it} = x_t)$, that is, by ignoring the heterogeneity entirely.
In the unobserved effects probit model, if $c_i$ is independent of $\{x_{i1}, \ldots, x_{iT}\}$ with a $Normal(0, \sigma^2_c)$ distribution, it can be show that

$$P(y_{it} = 1|x_{it}) = \Phi(x_{it}\beta_c),$$

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

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 $c_i$ and $x_i$, but it is restricted in some way.
- In a parametric setting, we specify a distribution for $D(c_i|x_{i1}, \ldots, x_{iT})$, 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

\[ D(c_i|x_{i1},...,x_{iT}) = D(c_i|\bar{x}_i), \]

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

• Often \( D(c_i|x_i) = D(c_i|\bar{x}_i) \) is used in conjunction with flexible parametric models for \( D(c_i|\bar{x}_i) \).
• 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 $D(c_i|x_i)$ to depend on individual-specific trends or measures of dispersion in $\{x_{it} : t = 1, \ldots, T\}$. 
Fixed Effects

• The label “fixed effects” is used in different ways.

1. The $c_i, i = 1, \ldots, N$ 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. $D(c_i|x_i)$ is unrestricted and we look for objective functions that do not depend on $c_i$ but still identify the population parameters. Leads to “conditional MLE” (CMLE) if we can find sufficient statistics $s_i$ such that

$$D(y_{i1}, \ldots, y_{iT}|x_i, c_i, s_i) = D(y_{i1}, \ldots, y_{iT}|x_i, s_i)$$

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 $c_i$. 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

\[ P(y_{it} = 1|x_{it}, c_i) = \Phi(x_{it}\beta + c_i), \ t = 1, \ldots, T. \]

- Strict exogeneity conditional on \( c_i \):

\[ P(y_{it} = 1|x_{i1}, \ldots, x_{iT}, c_i) = P(y_{it} = 1|x_{it}, c_i), \ t = 1, \ldots, T. \]

- Conditional independence [where we condition on \( x_i = (x_{i1}, \ldots, x_{iT}) \) and \( c_i \)]:

\[ D(y_{i1}, \ldots, y_{iT}|x_i, c_i) = D(y_{i1}|x_i, c_i) \cdot \cdots \cdot D(y_{iT}|x_i, c_i) \]
• Model for $D(c_i|\mathbf{x}_i)$:

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

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

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

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

$$u_{it}|(\mathbf{x}_{it}, c_i) \sim \text{Normal}(0, 1)$$

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

$$\{u_{it} : t = 1, \ldots, T\} \text{ independent}$$
• Can include time dummies in $x_{it}$ (but omit from $\tilde{x}_i$). Can also include time-constant elements as extra controls.

• If $\xi = 0$, get the traditional random effects probit model.

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

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

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

• 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 $x_t$, too, or other specific values.
As shown in Wooldridge (2011, Economics Letters), the unconditional heterogeneity distribution is consistently estimated as

\[
\hat{g}(c) = N^{-1} \sum_{i=1}^{N} \phi[(c - \hat{\psi} - \bar{x}_i \hat{\xi})/\hat{\sigma}_a]/\hat{\sigma}_a, \ c \in \mathbb{R}
\]
• The APEs are identified from the average structural function, easily estimated as

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

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

• Take derivatives and changes with respect to $$x_t$$. Can further average out across $$x_{i t}$$ 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
\[
P(y_{it} = 1|x_i) = P(y_{it} = 1|x_{it}, \bar{x}_i) = \Phi(x_{it}\beta + \psi_a + \bar{x}_i\xi_a).
\]
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

$$E(y_{it}|x_{it}, c_i) = \Phi(x_{it}\beta + c_i).$$

When the heterogeneity is integrated out,

$$E(y_{it}|x_{it}, \bar{x}_i) = \Phi(x_{it}\beta_a + \psi_a + \bar{x}_i\xi_a).$$
• Now exploit that the Bernoulli distribution is in the linear exponential family. Pooled “probit” is 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

<table>
<thead>
<tr>
<th>variable name</th>
<th>storage</th>
<th>display</th>
<th>value</th>
<th>label</th>
<th>variable label</th>
</tr>
</thead>
<tbody>
<tr>
<td>lfp</td>
<td>byte</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>=1 if in labor force</td>
</tr>
<tr>
<td>kids</td>
<td>byte</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>number children &lt; 18</td>
</tr>
<tr>
<td>hinc</td>
<td>float</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td>husband’s monthly income, $</td>
</tr>
</tbody>
</table>

. tab period

<table>
<thead>
<tr>
<th>months long</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5,663</td>
<td>20.00</td>
<td>20.00</td>
</tr>
<tr>
<td>2</td>
<td>5,663</td>
<td>20.00</td>
<td>40.00</td>
</tr>
<tr>
<td>3</td>
<td>5,663</td>
<td>20.00</td>
<td>60.00</td>
</tr>
<tr>
<td>4</td>
<td>5,663</td>
<td>20.00</td>
<td>80.00</td>
</tr>
<tr>
<td>5</td>
<td>5,663</td>
<td>20.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total | 28,315 | 100.00 |

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

. egen lhincbar = mean(lhinc), by(id)
```
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>FE</td>
<td>−.0389</td>
<td>−.199</td>
<td>−.0660</td>
<td>−.117</td>
<td>−.0389</td>
<td>−.317</td>
<td>−.0403</td>
</tr>
<tr>
<td>Probit</td>
<td>Pooled MLE</td>
<td>(.0092)</td>
<td>(.015)</td>
<td>(.0048)</td>
<td>(.027)</td>
<td>(.0085)</td>
<td>(.062)</td>
<td>(.0104)</td>
</tr>
<tr>
<td>CRE Probit</td>
<td>Pooled MLE</td>
<td>−.0089</td>
<td>−.211</td>
<td>−.0701</td>
<td>−.029</td>
<td>−.0095</td>
<td>−.078</td>
<td>−.0099</td>
</tr>
<tr>
<td>CRE Probit</td>
<td>MLE</td>
<td>(.0046)</td>
<td>(.024)</td>
<td>(.0079)</td>
<td>(.014)</td>
<td>(.0048)</td>
<td>(.041)</td>
<td>(.0055)</td>
</tr>
<tr>
<td>FE Logit</td>
<td>MLE</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

| kids       | —     | —     | —     | −.086 | —     | −.210 | —     | —     |
| lhinc      | —     | —     | —     | (.031)| —     | (.071)| —     | —     |
| kids       | —     | —     | —     | (.250)| —     | (.646)| —     | —     |
| lhinc      | —     | —     | —     | (.035)| —     | (.079)| —     | —     |

48
**Linear model by FE:**

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

**Fixed-effects (within) regression**

|                    | Coef.  | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------------------|--------|-----------|-------|-------|----------------------|
| **lfp**            |        |           |       |       |                      |
| kids               | -0.039 | 0.009     | -4.24 | 0.000 | -0.057 to -0.020     |
| lhinc              | -0.009 | 0.005     | -1.95 | 0.052 | -0.018 to 0.000      |
| per2               | -0.004 | 0.003     | -1.26 | 0.208 | -0.011 to 0.003      |
| per3               | -0.011 | 0.004     | -2.60 | 0.009 | -0.019 to -0.003     |
| per4               | -0.012 | 0.004     | -2.74 | 0.006 | -0.021 to -0.003     |
| per5               | -0.017 | 0.005     | -3.64 | 0.000 | -0.027 to -0.007     |
| _cons              | 0.809  | 0.038     | 21.56 | 0.000 | 0.735 to 0.883       |
| **sigma_u**        | 0.422  |           |       |       |                      |
| **sigma_e**        | 0.213  |           |       |       |                      |
| **rho**            | 0.796  |           |       |       | (fraction of variance due to u_i) |
```

Number of obs = 28315
Number of groups = 5663

(Std. Err. adjusted for 5663 clusters in id)
. * 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
Prob > chi2 = 0.0000

Log likelihood = -2003.4184

------------------------------------------------------------------------------
lfp | Coef. Std. Err. z P>|z| [95% Conf. Interval]
------------- ----------------------------- ----------------------------- -----------------------------
kids | -0.6438386 .1247828 -5.16 0.000 -.8884084 -.3992688
lhinc | -0.1842911 .0826019 -2.23 0.026 -.3461878 -.0223943
per2 | -0.0928039 .0889937 -1.04 0.297 -.2672283 .0816205
per3 | -0.2247989 .0887976 -2.53 0.011 -.398839 -.0507587
per4 | -0.2479323 .0888953 -2.79 0.005 -.422164 -.0737006
per5 | -0.3563745 .0888354 -4.01 0.000 -.5304886 -.1822604
------------------------------------------------------------------------------

. di 644/184
3.5

. di 389/89
4.3707865

50
. * CRE probit:

. xttprobit 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

Log likelihood = -8990.0898
Wald chi2(12) = 824.11
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  -.0519705
per3 | -.0954482  .0439688 -2.17 0.030  -.1816253  -.0092713
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  64.58511  1.78896
-----------------------------------------------------------------------------
sigma_u |  2.383059  .0317277  2.321679  2.446063
r homophobicu |  .8502764  .0038999  .8435102  .8567997
-----------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) = 15e+04 Prob >= chibar2 = 0.000

51
. 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  Coef. Std. Err. z P>|z| [95% Conf. Interval]
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>lfp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kids</td>
<td>-.1174</td>
<td>.02697</td>
<td>-4.35</td>
<td>0.000 -1.702435 -0.0645064</td>
</tr>
<tr>
<td>lhinc</td>
<td>-.02881</td>
<td>.01434</td>
<td>-2.01</td>
<td>0.045 -0.0569234 -0.0006961</td>
</tr>
<tr>
<td>kidsbar</td>
<td>-.08569</td>
<td>.03119</td>
<td>-2.75</td>
<td>0.006 -0.146814 -0.0245685</td>
</tr>
<tr>
<td>lhincbar</td>
<td>-.25018</td>
<td>.0353</td>
<td>-7.09</td>
<td>0.000 -0.3193466 -0.1810097</td>
</tr>
<tr>
<td>educ</td>
<td>.08414</td>
<td>.00673</td>
<td>12.50</td>
<td>0.000 .0709428 .0973248</td>
</tr>
<tr>
<td>black</td>
<td>.20307</td>
<td>.06639</td>
<td>3.06</td>
<td>0.002 .0729359 .3331976</td>
</tr>
<tr>
<td>age</td>
<td>.15164</td>
<td>.01283</td>
<td>12.15</td>
<td>0.000 .127176 .1761089</td>
</tr>
<tr>
<td>agesq</td>
<td>-.0002</td>
<td>.00015</td>
<td>-13.31</td>
<td>0.000 -.0023717 -.0017628</td>
</tr>
<tr>
<td>per2</td>
<td>-0.1357</td>
<td>.01037</td>
<td>-1.31</td>
<td>0.191 -.0339051 .0067648</td>
</tr>
<tr>
<td>per3</td>
<td>-.0332</td>
<td>.01272</td>
<td>-2.61</td>
<td>0.009 -.0581293 -.008269</td>
</tr>
<tr>
<td>per4</td>
<td>-.03903</td>
<td>.01362</td>
<td>-2.86</td>
<td>0.004 -.0657351 -.0123284</td>
</tr>
<tr>
<td>per5</td>
<td>-.05524</td>
<td>.01461</td>
<td>-3.78</td>
<td>0.000 -.0838711 -.0266139</td>
</tr>
<tr>
<td>cons</td>
<td>-.72605</td>
<td>.28369</td>
<td>-2.56</td>
<td>0.010 -1.282095 -.1700173</td>
</tr>
</tbody>
</table>
------------------------------------------------------------------------------
. drop xdhat xdhata
. predict xdhat, xb
. gen scale = normden(xdhat)
. sum scale

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>scale</td>
<td>28315</td>
<td>.3310079</td>
<td>.057301</td>
<td>.0694435</td>
<td>.3989423</td>
</tr>
</tbody>
</table>

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

. di .331*(-.02881)
-.00953611
Average marginal effects on Prob(lfp==1) after probit

|        | 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 - -0.0023  |
| 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)

|        | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|--------|-----------|-------|------|----------------------|
| lfp    |        |           |       |      |                      |
| kids   | -0.199 | 0.015     | -12.99| 0.00 | -0.229 to -0.169     |
| lhinc  | -0.211 | 0.024     | -8.69 | 0.00 | -0.259 to -0.163     |
| educ   | 0.079  | 0.006     | 12.17 | 0.00 | 0.067 to 0.092       |
| black  | 0.220  | 0.066     | 3.35  | 0.00 | 0.187 to 0.253       |
| age    | 0.145  | 0.012     | 11.86 | 0.00 | 0.121 to 0.169       |
| agesq  | -0.002 | 0.000     | -13.08| 0.00 | -0.023 to -0.001     |
| per2   | -0.012 | 0.010     | -1.19 | 0.23 | -0.033 to 0.009      |
| per3   | -0.032 | 0.013     | -2.55 | 0.01 | -0.057 to -0.008     |
| per4   | -0.046 | 0.014     | -3.38 | 0.00 | -0.073 to -0.019     |
| per5   | -0.058 | 0.015     | -3.95 | 0.00 | -0.087 to -0.029     |
| _cons  | -1.064 | 0.262     | -4.06 | 0.00 | -1.577 to -0.551     |
. margeff

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

|      | Coef.   | Std. Err. |   z  |   P>|z|  | [95% Conf. Interval] |
|------|---------|-----------|------|-------|----------------------|
| kids | -0.0660184 | 0.0049233 | -13.41 | 0.000   | -0.0756678 - -0.056369 |
| lhinc| -0.070054  | 0.0079819 | -8.78  | 0.000   | -0.0856981 - -0.0544099 |
| educ | 0.0264473  | 0.0021119 | 12.52  | 0.000   | 0.0223082 - 0.0305865  |
| black| 0.0698835  | 0.0197251 | 3.54   | 0.000   | 0.031223 - 0.108544   |
| age  | 0.0480966  | 0.0039216 | 12.26  | 0.000   | 0.0404105 - 0.0557828  |
| agesq| -0.0006609 | 0.000486  | -13.60 | 0.000   | -0.0007561 - -0.0005656 |
| per2 | -0.0041304 | 0.004828  | -1.19  | 0.236   | -0.0109565 - 0.0026957 |
| per3 | -0.010839  | 0.0042694 | -2.54  | 0.011   | -0.0192069 - -0.0024712 |
| per4 | -0.0153921 | 0.0045809 | -3.36  | 0.001   | -0.0243705 - -0.0064137 |
| per5 | -0.0193224 | 0.0049309 | -3.92  | 0.000   | -0.0289867 - -0.0096581 |

* 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 | Coef. Std. Err. z P>|z| [95% Conf. Interval] |
|---------------------|-------------|----------------|----------|-----------------|-------------------|
| bs_1 | -0.038852 .0085179 -4.56 0.000 -.0555469 -.0221572 |
| bs_2 | -0.0095363 .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.
11. end

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

59
probit_boot
(running probit_boot on estimation sample)

Bootstrap replications (500)
------------- 1 ------------- 2 ------------- 3 ------------- 4 ------------- 5

.................................................. 50

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

Bootstrap results

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

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Bootstrap</th>
<th>Normal-based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
<td>z</td>
</tr>
<tr>
<td>bs_1</td>
<td>-.0660184</td>
<td>.0047824</td>
<td>-13.80</td>
</tr>
<tr>
<td>bs_2</td>
<td>-.070054</td>
<td>.0078839</td>
<td>-8.89</td>
</tr>
</tbody>
</table>

. program drop probit_boot
end of do-file
• Many useful embellishments. For example, we can allow
\[ c_i | x_i \sim Normal[\psi + \bar{x}_i \xi, \sigma_a^2 \exp(\bar{x}_i \xi)], \]
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

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

still straightforward. For continuous \(x_{ij}\),

\[
\hat{APE}_j(x_t) = \hat{\beta}_j \left\{ N^{-1} \sum_{i=1}^{N} (1 + \hat{\sigma}_a^2 \exp(\bar{x}_i \hat{\zeta}))^{-1/2} \right. \\
\left. \cdot \phi[(1 + \hat{\sigma}_a^2 \exp(\bar{x}_i \hat{\zeta}))^{-1/2} (x_{i}\hat{\beta}_a + \hat{\psi}_a + \bar{x}_i \hat{\xi}_a)] \right\}
\]

• See Wooldridge (2010, Chapter 15).
6. CRE Tobit Model

• Unobserved effects Tobit model for a corner at zero is

\[ y_{it} = \max(0, x_{it} \beta + c_i + u_{it}) \]

\[ D(u_{it}|x_{it}, c_i) = \text{Normal}(0, \sigma_u^2) \]

• Strict exogeneity conditional on \( c_i \):

\[ D(u_{it}|x_{it}, c_i) = D(u_{it}|x_{it}, c_i) \]

• Conditional independence: The \( \{u_{it} : t = 1, \ldots, T\} \) are independent.
• Model for $D(c_i|x_i)$:

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

• Joint MLE (conditional on $x_i$) is relatively straightforward. It is based on the joint distribution $D(y_{i1}, \ldots, y_{iT}|x_i)$. 
• 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 $\sigma_c^2$:

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

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

• Same estimate of heterogeneity distribution works, too.
• We can evaluate the partial effects of the Tobit function, 
\( m(x_t \hat{\beta} + c, \hat{\sigma}_u^2) \) 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 \( x_t \). For a continuous variable,

\[
\hat{\beta}_j \Phi[(x_t \hat{\beta} + c)/\hat{\sigma}_u]
\]

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

\[
\hat{ASF}(x_t) = N^{-1} \sum_{i=1}^{n} m(x_t \hat{\beta} + \hat{\psi} + \bar{x}_i \hat{\xi}, \hat{\sigma}_a^2 + \hat{\sigma}_u^2)
\]

where \( m(z, \sigma^2) \) is the mean function for a Tobit.
• Take derivatives and differences with respect to elements of $x_t$. Panel bootstrap for inference.

• For a continuous $x_{ij}$,

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

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

67
• If we drop the conditional independence assumption and allow and serial dependence in \( \{u_{it}\} \) then we only have the marginal distributions

\[
D(y_{it}|x_i) = D(y_{it}|x_{it}, \bar{x}_i) = \text{Tobit}(x_{it}\beta + \psi + \bar{x}_i\xi, \sigma_v^2)
\]

• So, we can apply pooled Tobit, ignoring the serial correlation, to estimate \( \beta, \psi, \xi, \) 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 \( \beta \) nor \( \sigma_u^2 \).
. use psid80_92

. des hours nwifeinc exper ch0_2 ch3_5 ch6_17

<table>
<thead>
<tr>
<th>variable</th>
<th>type</th>
<th>format</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>hours</td>
<td>int</td>
<td>%9.0g</td>
<td>annual work hours</td>
</tr>
<tr>
<td>nwifeinc</td>
<td>float</td>
<td>%9.0g</td>
<td>(faminc - wife’s labor income)/1000</td>
</tr>
<tr>
<td>exper</td>
<td>float</td>
<td>%9.0g</td>
<td>years of workforce experience</td>
</tr>
<tr>
<td>ch0_2</td>
<td>byte</td>
<td>%9.0g</td>
<td>number of children in FU, 0-2</td>
</tr>
<tr>
<td>ch3_5</td>
<td>byte</td>
<td>%9.0g</td>
<td>number of children in FU, 3-5</td>
</tr>
<tr>
<td>ch6_17</td>
<td>byte</td>
<td>%9.0g</td>
<td>number of children in FU, 6-17</td>
</tr>
</tbody>
</table>

. sum hours nwifeinc exper ch0_2 ch3_5 ch6_17

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>hours</td>
<td>11674</td>
<td>1130.995</td>
<td>888.2304</td>
<td>0</td>
<td>5168</td>
</tr>
<tr>
<td>nwifeinc</td>
<td>11674</td>
<td>34.22192</td>
<td>40.00195</td>
<td>-14.99792</td>
<td>1412.2</td>
</tr>
<tr>
<td>exper</td>
<td>11674</td>
<td>11.80465</td>
<td>7.743591</td>
<td>0</td>
<td>45.1315</td>
</tr>
<tr>
<td>ch0_2</td>
<td>11674</td>
<td>.1351722</td>
<td>.3700769</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>ch3_5</td>
<td>11674</td>
<td>.1776598</td>
<td>.42228</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>ch6_17</td>
<td>11674</td>
<td>.8222546</td>
<td>1.008326</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>
. tab year

<table>
<thead>
<tr>
<th></th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>898</td>
<td>7.69</td>
<td>7.69</td>
</tr>
<tr>
<td>81</td>
<td>898</td>
<td>7.69</td>
<td>15.38</td>
</tr>
<tr>
<td>82</td>
<td>898</td>
<td>7.69</td>
<td>23.08</td>
</tr>
<tr>
<td>83</td>
<td>898</td>
<td>7.69</td>
<td>30.77</td>
</tr>
<tr>
<td>84</td>
<td>898</td>
<td>7.69</td>
<td>38.46</td>
</tr>
<tr>
<td>85</td>
<td>898</td>
<td>7.69</td>
<td>46.15</td>
</tr>
<tr>
<td>86</td>
<td>898</td>
<td>7.69</td>
<td>53.85</td>
</tr>
<tr>
<td>87</td>
<td>898</td>
<td>7.69</td>
<td>61.54</td>
</tr>
<tr>
<td>88</td>
<td>898</td>
<td>7.69</td>
<td>69.23</td>
</tr>
<tr>
<td>89</td>
<td>898</td>
<td>7.69</td>
<td>76.92</td>
</tr>
<tr>
<td>90</td>
<td>898</td>
<td>7.69</td>
<td>84.62</td>
</tr>
<tr>
<td>91</td>
<td>898</td>
<td>7.69</td>
<td>92.31</td>
</tr>
<tr>
<td>92</td>
<td>898</td>
<td>7.69</td>
<td>100.00</td>
</tr>
</tbody>
</table>

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  Prob > F = 0.0000
(Std. Err. adjusted for 898 clusters in id)

|        | Coef. | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|--------|-------|-----------|------|-------|----------------------|
| hours  |       |           |      |       |                      |
* 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
Observations per group:
  min = 13
  avg = 13.0
  max = 13

Wald chi2(22) = 1501.20
Prob > chi2 = 0.0000

|                  | Coef.  | Std. Err. | z    | P>|z|   | [95% Conf. Interval] |
|------------------|--------|-----------|------|-------|---------------------|
| hours            |        |           |      |       |                     |
| nwifeinc         | -1.554228 | 0.3816927 | -4.07 | 0.000 | -2.302332 to -0.8061243 |
| ch0_2            | -472.088 | 23.03087  | -20.50 | 0.000 | -517.2277 to -426.9483 |
| ch3_5            | -329.3896 | 19.49411  | -16.90 | 0.000 | -367.5974 to -291.1819 |
| ch6_17           | -46.11619 | 10.89609  | -4.23  | 0.000 | -67.47213 to -24.76024 |
| marr             | -784.1809 | 155.0133  | -5.06  | 0.000 | -1088.001 to -480.3604 |
| y81              | -7.060588 | 31.52257  | -0.22  | 0.823 | -68.84369 to 54.72251 |
| y82              | -38.9034  | 31.70009  | -1.23  | 0.220 | -101.0344 to 23.22764 |
| y83              | -9.719573 | 31.68694  | -0.31  | 0.759 | -71.82483 to 52.38569 |
| y84              | 99.77618  | 31.61932  | 3.16   | 0.002 | 37.80345 to 161.7489 |
| y85              | 89.15912  | 31.7439   | 2.81   | 0.005 | 26.94222 to 151.376 |
| y86              | 82.60212  | 31.76385  | 2.60   | 0.009 | 20.34612 to 144.8581 |
| y87              | 48.59097  | 31.98439  | 1.52   | 0.129 | -14.09729 to 111.2792 |
| y88              | 53.52189  | 32.09804  | 1.67   | 0.095 | -9.389108 to 116.4329 |
| y89              | 68.69013  | 32.23667  | 2.13   | 0.033 | 5.507141 to 131.8728 |
| y90              | 71.2654   | 32.3657   | 2.20   | 0.028 | 7.8298 to 134.701 |
| y91              | 64.89096  | 32.48217  | 2.00   | 0.046 | 1.227067 to 128.5548 |
| y92              | 4.334129  | 32.82961  | 0.13   | 0.895 | -60.01072 to 68.67898 |
| nwifeincb        | -7.639696 | 0.6815067 | -11.21 | 0.000 | -8.975424 to -6.303967 |
| ch0_2b           | -143.4709 | 155.0915  | -0.93  | 0.355 | -447.4448 to 160.5029 |
| ch3_5b           | 531.2027  | 150.3881  | 3.53   | 0.000 | 236.4475 to 825.9578 |
| ch6_17b          | 5.854889  | 28.04159  | 0.21   | 0.835 | -49.10563 to 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)

73
```
-1.300698
.dl (0.837)*(-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

|                  | Coef.    | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|------------------|----------|-----------|-------|-------|----------------------|
| nwifeinc         | -1.796524 | .6073205  | -2.96 | 0.003 | -2.986975, -1.606744 |
| 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 | 103.5075, 103.8165   |
| y83              | 1544965   | 52.88423  | 2.11  | 0.034 | 8.181439, 215.3372   |
| y84              | 98.8203   | 53.02693  | 1.86  | 0.062 | 5.121366, 202.762    |
| y86              | 91.11739  | 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  -37.1646  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
Group variable (i): id
Number of obs  = 11674
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  48.93359  -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  -69.57433   55.03156
y84    | 104.3436  31.71544   3.29 0.001  42.18249  166.50472
y85    | 94.90622  31.82266   2.98 0.003  32.53496  157.27745
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  144.6783  
y89 |  81.55682  32.20542  2.53  0.011   18.43536  144.6783  
y90 |  85.75216  32.26838  2.66  0.008   22.50728  144.6783  
y91 |  80.93763  32.36379  2.50  0.012   17.50576  144.6783  
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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHIhata</td>
<td>898</td>
<td>0.8240658</td>
<td>0.0724009</td>
<td>0.3761031</td>
<td>0.9578886</td>
</tr>
</tbody>
</table>

* 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:

\[ E(y_{it}|x_{it}, c_i) = c_i \exp(x_{it}\beta) \]

where \( c_i \geq 0 \) is the unobserved effect and \( x_{it} \) would include 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 $c_i$:

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

• 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

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

where $a_i$ is independent of $x_i$ with unit mean. Then

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

• So, can use any common method and simply add $\bar{x}_i$ as a set of covariates. Can add time-constant covariates, too.

• Can easily test $H_0 : \xi = 0$. 
• Stata commands:

\[ \text{glm } y \text{ x1 ... xK x1bar ... xKbar, fam(poisson)} \]
\[ \text{cluster(id)} \]
\[ \text{xtgee } y \text{ x1 ... xK x1bar ... xKbar, fam(poisson)} \]
\[ \text{corr(uns) robust} \]
\[ \text{xtpoisson } y \text{ x1 ... xK x1bar ... xKbar, re} \]

• Pooled Poisson and GEE only use \( E(y_{it}|x_i) = \exp(\psi + x_{it}\beta + \bar{x}_i\xi) \).

The Poisson RE method requires that \( D(y_{it}|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 $c_i$:

$$E(y_{it}|x_i, c_i) = E(y_{it}|x_{it}, c_i) = m_i(x_{it}, c_i)$$

- But we do not want to use a parametric model for $D(c_i|x_i)$. Maybe we want to leave $m_i(\cdot, \cdot)$ unspecified, too.

- Altonji and Matzkin (2005, Econometrica) show how to identify the average structural function (and a local version) by using “exchangeability” assumptions on $D(c_i|x_i)$. 
• Leading exchangeability assumption:

\[ D(c_i|x_i) = D(c_i|\bar{x}_i) \]

• But \( D(c_i|x_i) \) might depend also on the sample variance-covariance matrix,

\[ S_i = (T - 1)^{-1} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)'(x_{it} - \bar{x}_i) \]
• Generally, let $w_i$ be a set of exchangeable functions of $\{x_{it}\}$ such that

\[
D(c_i|x_i) = D(c_i|w_i)
\]

• Under restrictions on the nature of $w_i$, the ASF can be identified from

\[
E(y_{it}|x_i, w_i) \equiv g_t(x_{it}, w_i).
\]

\[
ASF(x_t) = E_{w_i}[g_t(x_t, w_i)].
\]

• Practically, might implement using flexible parametric forms for $g_t(x_{it}, w_i)$. 
• For example, if \( w_i = \bar{x}_i \) and \( 0 \leq y_{it} \leq 1 \) we can just start with

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

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

• For a continuous variable \( x_{tj} \) the estimated APE is

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

• If the model were linear, the pooled OLS estimates of \( \beta \) and \( \delta \) would be the FE estimates.
For $y_{it} \geq 0$ can use

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

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 \( \{x_{it}\} \).

• 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 \((x_{it}, w_i)\) with \(w_i = \bar{x}_i\) the leading but not only case.