CRE METHODS FOR UNBALANCED PANELS

Correlated Random Effects Panel Data Models

IZA Summer School in Labor Economics
May 13-19, 2013
Jeffrey M. Wooldridge
Michigan State University

1. Introduction
2. Linear Model with Additive Heterogeneity
3. Linear Model with Correlated Random Slopes
4. A Modeling Approach for Nonlinear Models
5. Estimating Average Partial Effects
6. Application to Probit/Fractional Response
1. Introduction

• In linear model with additive heterogeneity, unbalanced panels cause no serious issues provided certain assumptions about selection hold.
• FE is more robust than RE in the sense that with RE selection must be assumed uncorrelated with heterogeneity as well as with idiosyncratic shocks. FE allows arbitrary correlation between selection and $c_i$. 
• For nonlinear models, RE easily adapts to unbalanced panels, but the selection assumptions are strong.

• Conditional MLE allows unbalanced panels when it applies (logit, Poisson).

• Treating the $c_i$ as parameters to estimate – so called “fixed effects” – does not require a balanced panel (but still has an incidental parameters problem with small T).
• Drawback to correlated RE approach compared with CMLE or FE: Not clear how to allow for unbalanced panels when we want heterogeneity to be correlated with covariates and selection.

• Want to be able to handle nonlinear CRE models.

• Without specifically modeling the selection rule – say, using a Heckman-type approach – all methods assume selection is independent of shocks. We can test this assumption.
2. Linear Model with Additive Heterogeneity

• We still draw a random sample from the cross section, with units indexed by \( i \). But now we may not observe a complete set of time series observations.

• Model this situation using a sequence of selection indicators, \( \{s_{i1}, \ldots, s_{iT}\} \), where \( s_{it} = 1 \) if time period \( t \) for unit \( i \) can be used in estimation. Usually this means that we observe all elements of \( (x_{it}, y_{it}) \).
• We only use an \((i, t)\) pair when a full set of data is observed, as happens when software is used for RE and FE on unbalanced panels. So we focus on complete-case estimators without imputation.

• We still write the population model for a random draw \(i\) as

\[
y_{it} = \mathbf{x}_{it} \boldsymbol{\beta} + c_i + u_{it}, \quad t = 1, \ldots, T,
\]

where \(\mathbf{x}_{it}\) can generally include a fully set of time dummies, or other aggregate time variables.
• Along with a rank condition, a sufficient (but not necessary) condition for consistency of FE on the unbalanced panel is

$$E(u_{it}|x_i, c_i, s_i) = 0, \ t = 1, \ldots, T,$$

where \(x_i = (x_{i1}, x_{i2}, \ldots, x_{iT})\) and \(s_i = (s_{i1}, s_{i2}, \ldots, s_{iT})\).

• Allows selection in any time period to be correlated with \((x_i, c_i)\), but selection in all time periods must be unrelated to the idiosyncratic shocks.
The time-demeaned data now uses different time periods for different $i$. Let

$$
\hat{y}_{it} = y_{it} - T_i^{-1} \sum_{r=1}^{T} s_{ir} y_{ir} = y_{it} - \bar{y}_i
$$

$$
\hat{x}_{it} = x_{it} - T_i^{-1} \sum_{r=1}^{T} s_{ir} x_{ir} = x_{it} - \bar{x}_i
$$

where $T_i = \sum_{r=1}^{T} s_{ir}$ is random.
• The FE estimator is then

$$\hat{\beta} = \left( N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} S_{it} \bar{x}_{it}' \bar{x}_{it} \right)^{-1} \left( N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} S_{it} \bar{x}_{it}' \bar{y}_{it} \right)$$

$$= \beta + \left( N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} S_{it} \bar{x}_{it}' \bar{x}_{it} \right)^{-1} \left( N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} S_{it} \bar{x}_{it}' u_{it} \right)$$

• Consistency (fixed $T$, $N \to \infty$) follows if

$$\text{rank} \left[ \sum_{t=1}^{T} E(s_{it} \bar{x}_{it}' \bar{x}_{it}) \right] = K$$

$$E(s_{it} \bar{x}_{it}' u_{it}) = 0, \ t = 1, \ldots, T.$$
• In the balanced case we know that if we estimate the equation
\[ y_{it} = x_{it} \beta + \psi + \bar{x}_i \xi + v_{it} \]
by pooled OLS or RE we get the FE estimate for \( \beta \).

• Conveniently, the same result holds for unbalanced case, with a
  caveat: the time averages are defined only for the complete-case
  observations.
• In other words, estimate the equation

\[ y_{it} = x_{it} \beta + \psi + \bar{x}_i \xi + \nu_{it} \]

by pooled OLS using the \( s_{it} = 1 \) observations. The coefficient vector \( \hat{\beta} \) is identical to the fixed effects.
• Unlike in the balanced case, any aggregate time variables, including time dummies, should be part of $x_{it}$, and their time averages must be included in $\bar{x}_i$ to get the FE estimates on the other variables.

• Wooldridge (2010, working paper) shows a more general result. Recall that the RE estimator can be obtained from a pooled OLS regression. On an unbalanced panel, it is

$$y_{it} - \theta_i \bar{y}_i \text{ on } x_{it} - \theta_i \bar{x}_i, \ (1 - \theta_i)\bar{z}_i, \ (1 - \theta_i)z_i \text{ if } s_{it} = 1,$$

where

$$\theta_i = 1 - \left[\frac{\sigma_u^2}{(\sigma_u^2 + T_i\sigma_c^2)}\right]^{1/2}$$

varies because $T_i$ varies.
**Algebraic Fact:** Let $\tilde{\beta}$ be the vector $(K \times 1)$ of coefficients on $x_{it} - \theta_i \bar{x}_i$ in the POLS regression above. Then $\tilde{\beta} = \hat{\beta}_{FE}$, the fixed effects estimate on the unbalanced panel.

- Note that $z_i$ can contain any time-constant variables, including functions of $T_i$, or interactions of the form $T_i \cdot \bar{x}_i$, or allow a different set of coefficients on $\bar{x}_i$ for each different $T_i$. The coefficients on $\bar{x}_i$ may changed widely across $T_i$, yet the estimate on $x_{it}$ is still the FE estimate.
• The algebraic equivalence still justifies the robust, regression-based Hausman test.

• Write a model with time-constant variables $z_i$ as

$$y_{it} = x_{it} \beta + z_i \gamma + c_i + u_{it}, \quad t = 1, \ldots, T,$$

where $z_i$ includes a constant.

• Use the Mundlak equation

$$y_{it} = x_{it} \beta + \bar{x}_i \xi + z_i \gamma + a_i + u_{it}$$

and estimate this by RE.
The regression based Hausman test is just a (robust) Wald test of $H_0 : \xi = 0$ after RE estimation of the augmented equation.

Remember that FE (and RE) assume that selection in any time period is not correlated with $u_{it}$. Might worry that a shock today affects being in the sample in the future.

Wooldridge (2010, MIT Press): Using FE on the unbalanced panel, estimate the equation

$$y_{it} = x_{it}\beta + \eta s_{i,t+1} + c_i + u_{it}, \quad t = 1, \ldots, T - 1$$

and test $H_0 : \eta = 0$ using a robust test.
• If we use RE we can (and should) test if $c_i$ is correlated with functions of $T_i$. For example, define dummies

$$d_{ir} = 1[T_i = r], r = 1, \ldots, T - 1$$

and add these to the usual RE estimation.

• Significance of $d_i = (d_{i1}, \ldots, d_{i,T-1})$ casts serious doubt on the usual RE analysis. FE does not care.
• We can jointly test the assumption that

\[ E(c_i|x_{i1}, \ldots, x_{iT}, s_{i1}, \ldots, s_{iT}) = E(c_i) \]

by estimating the equation

\[ y_{it} = x_{it}\beta + \psi + \bar{x}_i\xi + d_i\gamma + v_{it} \]

by RE using the unbalanced panel and jointly testing \( H_0 : \xi = 0, \gamma = 0 \).
3. Linear Model with Correlated Random Slopes

- Now consider a model with all slopes heterogeneous:

\[ E(y_{it}|x_i, a_i, b_i) = a_i + x_{it}b_i, \]

so, in the population, \( \{x_{it} : t = 1, \ldots, T\} \) is strictly exogenous conditional on \((a_i, b_i)\).

- Write \( a_i = \alpha + c_i \), \( b_i = \beta + d_i \) and

\[ y_{it} = \alpha + x_{it}\beta + c_i + x_{it}d_i + u_{it} \]

where \( E(u_{it}|x_i, a_i, b_i) = E(u_{it}|x_i, c_i, d_i) = 0 \) for all \( t \).
• Selection may be related to \((x_i, a_i, b_i)\) but not the idiosyncratic shocks:

\[
E(y_{it}|x_i, a_i, b_i, s_i) = E(y_{it}|x_i, a_i, b_i)
\]

or

\[
E(u_{it}|x_i, a_i, b_i, s_i) = 0, \ t = 1, \ldots, T.
\]

• With enough time periods, we could obtain \(\hat{a}_i, \hat{b}_i\) ("fixed effects") and then average these. But need \(T_i > K + 1\).
• Unfortunately, if selection is correlated with $b_i$, there are no intuitive robustness results for the usual FE estimator.

• How might we use a CRE approach?
• Multiply through by selection:

\[ s_{it} y_{it} = s_{it} \alpha + s_{it} x_{it} \beta + s_{it} c_i + s_{it} x_{it} d_i + s_{it} u_{it} \]

• Because we only use observations with \( s_{it} = 1 \), we handle the heterogeneity by conditioning on the entire history of selection and the values of the covariates if selected. That is, \( \{ (s_{it}, s_{it} x_{it}) : t = 1, \ldots, T \} \). If \( s_{it} = 0 \) the observation is not used; if \( s_{it} = 1 \), the observation is used, and we observe \( x_{it} \).
• If we condition on the larger information set 
\{(x_{i1}, s_{i1}), (x_{i2}, s_{i2}), \ldots, (x_{iT}, s_{iT})\} and heterogeneity depends only on the covariates, we would be left with an equation that is not estimable unless the covariates are always observed.
• Write $h_i \equiv \{h_{it} : t = 1, \ldots, T\} \equiv \{(s_{it}, s_{it}x_{it}) : t = 1, \ldots, T\}$. Then, extending Mundlak (1978) and Chamberlain (1982), we work with

$$E(s_{it}y_{it}|h_i) = s_{it}\alpha + s_{it}x_{it}\beta + s_{it}E(c_i|h_i) + s_{it}x_{it}E(d_i|h_i)$$

and then make assumptions concerning $E(c_i|h_i)$ and $E(d_i|h_i)$.

• Alternatively, we could eliminate $c_i$ using the within transformation, and focus just on $E(d_i|h_i)$. 
• A simple approach is to model the expectations as exchangeable functions of $\{h_{it} : t = 1, \ldots, T\}$ – extending the balanced case considered by Altonji and Matzkin (2005).

• From the model with constant slopes, natural to start with

$$w_i \equiv (T_i, \bar{x}_i).$$

• Fairly natural extension of Mundlak (1978).
• A flexible specification with $g_{ir} \equiv 1[T_i = r]$:

$$E(c_i|T_i, \bar{x}_i) = \sum_{r=1}^{T} \psi_r (g_{ir} - \rho_r) + \sum_{r=1}^{T} g_{ir} \cdot (\bar{x}_i - \mu_r)\xi_r$$

$$E(d_i|T_i, \bar{x}_i) = \sum_{r=1}^{T} (g_{ir} - \rho_r)k_r + \sum_{r=1}^{T} g_{ir} \cdot [(\bar{x}_i - \mu_r) \otimes I_K]\eta_r,$$

where

$$\rho_r = P(T_i = r) = E\{1[T_i = r]\}$$

$$\mu_r \equiv E(\bar{x}_i|T_i = r)$$
• As a practical matter, the formulation above is identical to running separate regressions for each $T_i$:

$$y_{it} \text{ on } 1, \ x_{it}, \ \bar{x}_i, \ (\bar{x}_i - \hat{\mu}_r) \otimes x_{it}, \ \text{for } s_{it} = 1$$

where $\hat{\mu}_r = N_r^{-1} \left( \sum_{i=1}^{N} 1[T_i = r] \bar{x}_i \right)$ and $N_r$ is the number of observations with $T_i = r$. The coefficient on $x_{it}, \hat{\beta}_r,$ is the APE given $T_i = r$. We can average these across $r$ to obtain the overall APE.
• Notice that we lose the $T_i = 1$ observations – just as in a standard fixed effects estimation.

• We could use grouping based on $T_i$ so that the APEs include the entire population.
• In any formulation, including the basic FE estimation, have a simple test for dynamic selection bias – that is, for the null

\[ E(y_{it}|x_i, a_i, b_i, s_i) = E(y_{it}|x_i, a_i, b_i). \]

Assumes that our model for \( E(c_i, d_i | \{(s_{it}, s_{it}x_{it}) : t = 1, \ldots, T\} \) is correctly specified. Then, no other functions of \( \{(s_{it}, s_{it}x_{it}) : t = 1, \ldots, T\} \) should appear in \( E(s_i y_{it}|h_i) \).
• Assuming we have some $T_i$ with $T_i \geq 3$, add as extra regressors at time $t$ the variables $(s_{i,t+1}, s_{i,t+1}x_{i,t+1})$. We can compute a fully robust (to serial correlation and heteroskedasticity) Wald test of joint significance. Produces a joint test of ignorable selection and strictly exogenous covariates.
4. A Modeling Approach for Nonlinear Models

• Interested in the conditional distribution

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

or maybe just an expectation.

• Assume strictly exogenous covariates conditional on \( c_i \) and ignorable selection:

\[ D(y_{it} | x_i, c_i, s_i) = D(y_{it} | x_{it}, c_i), \; t = 1, \ldots, T. \]
• Focus on methods that only exploit assumptions about the marginal distributions, $D(y_{it}|x_{it}, c_i)$, not joint distributions. Average partial effects are generally identified without restricting conditional dependence.

• Let $g_t(y_t|x_{it}, c_i; \gamma)$ be a parametric density conditional on $(x_{it}, c_i)$.

• As with the linear model, specify models for

$$ D(c_i|(s_{it}, s_{it}x_{it}) : t = 1, \ldots, T). $$

• Let $w_i$ be a vector of known functions of $\{(s_{it}, s_{it}x_{it}) : t = 1, \ldots, T\}$ that act as sufficient statistics, so that

$$ D(c_i|(s_{it}, s_{it}x_{it}) : t = 1, \ldots, T) = D(c_i|w_i) $$
• Because $D(y_{it}|x_{it},c_i,s_{it} = 1) = D(y_{it}|x_{it},c_i)$ and we have a density for the latter, we can obtain the density given $(x_{it},w_i,s_{it} = 1)$ as

$$f_t(y_t|x_{it},w_i;\gamma,\delta) = \int_{\mathbb{R}^M} g_t(y_t|x_{it},c;\gamma) h(c|w_i;\delta) dc$$

where $h(c|w_i;\delta)$ is a parametric density for $D(c_i|w_i)$ and $M$ is the dimension of $c_i$.

• The partial log-likelihood function for the entire sample is

$$\sum_{i=1}^{N} \sum_{t=1}^{T} s_{it} \log[f_t(y_{it}|x_{it},w_i;\gamma,\delta)].$$

• Need a robust sandwich estimator for serial correlation (and maybe other misspecifications).
• Same basic arguments carry through if we focus on estimating conditional means. We start with $E(y_{it}|x_{it}, c_i) = m_t(x_{it}, c_i)$ and obtain $E(y_{it}|x_{it}, w_i)$ by integrating $m_t(x_{it}, c_i)$ with respect to the density of $c_i$ given $w_i$ (which, again, is valid for the mean when $s_{it} = 1$). Or, we just assert a model for $E(y_{it}|x_{it}, w_i)$.

• Can then use a host of quasi-log likelihoods on the selected sample, including the Bernoulli for fractional responses, the gamma for nonnegative (continuous) responses, and the Poisson for nonnegative (count) responses.
5. Estimating Average Partial Effects

- In most nonlinear models, the parameters $\gamma$ appearing in $f_t(y_t|x_t,c;\gamma)$ provide only part of the story for the effect of $x_t$ on $y_t$. Using pooled methods we often cannot fully identify $\gamma$ or the distribution of $c_i$.
- Even in the unbalanced case we can use the Blundell and Powell (2003) approach provided a vector $w_i$ suitably proxies for correlation between $c_i$ and $\{(s_{it},s_{it}x_{it}) : t = 1, \ldots, T\}$. 
Let $q_t(x_t, w; \theta)$ denote the mean associated with $f_t(y_t|x_t, w; \theta)$. Then

$$ASF(x_t) = E_{w_i}[q_t(x_t, w_i; \theta)],$$

that is, we can obtain the ASF by averaging out the observed vector of sufficient statistics, $w_i$, from $E(y_{it}|x_t, w_i, s_{it} = 1)$ rather than averaging out $c_i$ from $E(y_{it}|x_t, c_i)$.

• Consistent estimation is then easy:

$$\hat{ASF}(x_t) = N^{-1} \sum_{i=1}^{N} q_t(x_t, w_i; \hat{\theta})$$

• Panel bootstrap is still justified with unbalanced panel.
• Tricky to estimate an overall average partial effect. With a balanced panel, it is natural to average $\widehat{APE}_{tj}(x_{it}, w_i)$ across the distribution of $(x_{it}, w_i)$, and then possibly across $t$, too. But if selection $s_{it}$ depends on $x_{it}$, averaging across the selected sample does not consistently estimate $E_{(x_{it}, w_i)}[APE_{tj}(x_{it}, w_i)]$.

• Presumably we still have an idea of useful values to plug in for $x_t$. 
6. Application to Probit/Fractional Response

- The probit model, either for true binary responses or fractional responses, is natural for applying the previous approach. The model is

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

where \( x_{it} \) can include time dummies or other aggregate time variables.

- Once we specify \( P(y_{it} = 1|x_{it}, c_i) \) and assume that selection is conditionally ignorable for all \( t \), that is,

\[ P(y_{it} = 1|x_{it}, c_i, s_i) = P(y_{it} = 1|x_{it}, c_i), \]

all that is left is to specify a model for \( D(c_i|w_i) \) for suitably chosen functions \( w_i \) of \( \{s_{it}, s_{it}x_{it}\} : t = 1, \ldots, T\).
• A specification linear in $\bar{x}_i$ but with intercept and slopes different for each $T_i$ is

$$E(c_i|w_i) = \sum_{r=1}^{T} \psi_r 1[T_i = r] + \sum_{r=1}^{T} 1[T_i = r] \cdot \bar{x}_i \xi_r$$

• At a minimum, should let the variance of $c_i$ change with $T_i$:

$$Var(c_i|w_i) = \exp\left(\tau + \sum_{r=1}^{T-1} 1[T_i = r]\omega_r\right)$$
• If we also maintain that $D(c_i|\mathbf{w}_i)$ is normal, then we obtain the following response probability for $s_{it} = 1$ (with normalization that does not affect estimation of APEs):

$$P(y_{it} = 1|x_{it}, \mathbf{w}_i, s_{it} = 1) = \Phi \left[ \frac{\mathbf{x}_{it}\beta + \sum_{r=1}^{T} \psi_r g_{ir} + \sum_{r=1}^{T} g_{ir} \cdot \bar{x}_i \xi_r}{\exp \left( \sum_{r=2}^{T} g_{ir} \omega_r \right)^{1/2}} \right]$$

so that the denominator is unity when all $\omega_r$ are zero. (Recall $g_{ir} = 1[T_i = r]$.)

• No difficulty in adding $g_{ir} \cdot \bar{x}_i$ for $r = 1, \ldots, T$ to the variance function.
• Computational bonus: Estimable by so-called “heteroskedastic probit” software, where the explanatory variables at time $t$ are $(1, x_{it}, g_{i1}, \ldots, g_{iT}, g_{i1} \cdot \bar{x}_i, \ldots, g_{iT} \cdot \bar{x}_i)$ and the explanatory variables in the variance are simply the dummy variables $(g_{i2}, \ldots, g_{iT})$, or also add $g_{i1} \cdot \bar{x}_i, \ldots, g_{iT} \cdot \bar{x}_i$. 
• The APEs are easy to obtain from the estimated ASF:

\[
\frac{\widehat{ASF}(x_t)}{N^{-1}} = \sum_{i=1}^{N} \Phi \left[ \frac{x_i \hat{\beta} + \sum_{r=1}^{T} \hat{\psi}_r g_{ir} + \sum_{r=1}^{T} g_{ir} \cdot \bar{x}_i \xi_r}{\exp \left( \sum_{r=2}^{T} g_{ir} \hat{\omega}_r \right)^{1/2}} \right]
\]

where the coefficients with "^" are from the pooled heteroskedastic probit estimation.

• The functions of \((T_i, \bar{x}_i)\) are averaged out, leaving the result a function of \(x_t\). If, say, \(x_{ij}\) is continuous, its APE is estimated as

\[
\frac{\widehat{APE}_{ij}(x_t)}{N^{-1}} = \hat{\beta}_j \left\{ \sum_{i=1}^{N} \phi \left[ \frac{x_i \hat{\beta} + \sum_{r=1}^{T} \hat{\psi}_r g_{ir} + \sum_{r=1}^{T} g_{ir} \cdot \bar{x}_i \xi_r}{\exp \left( \sum_{r=2}^{T} g_{ir} \hat{\omega}_r \right)^{1/2}} \right] \right\}
\]
• The above procedure applies, without change, if $y_{it}$ is a fractional response; that is, $0 \leq y_{it} \leq 1$. The original model is $E(y_{it}|x_{it}, c_i) = \Phi(x_{it}\beta + c_i)$ and APEs are on the mean response.

• Could allow $\hat{\beta}$ to change with each $T_i$ (losing $T_i = 1$). Then, just estimate an APE for each $T_i \geq 2$ and average the results. Just a separate probit with explanatory variables $(1, x_{it}, \bar{x}_i)$.

• Or, allow heteroskedasticity as a function of $\bar{x}_i$, with ASF

$$\hat{ASF}(x_t) = N^{-1} \sum_{r=2}^{T} \sum_{i=1}^{N} g_{ir} \Phi \left[ \frac{x_{it}\hat{\beta}_r + \hat{\psi}_r + \bar{x}_i\hat{\xi}_r}{\exp(\bar{x}_i\hat{\lambda}_r/2)} \right].$$
• Same ideas apply to ordered probit and even multinomial logit, as well as Tobit and count data models.

• With count data, Poisson FE estimator has very nice robustness properties, including for sample selection –

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

is sufficient – but only for the scalar, additive heterogeneity case.
• Scope for CRE approach in more complicated models, such as

\[ E(y_{it}|x_i, c_i) = \exp(a_i + x_{it}b_i). \]

• CMLE approach makes no sense with all coefficients heterogeneous (and it is not clear a CMLE exists, anyway).

• FE approach can be implemented with lots of large \( T_i \) \( (T_i > K + 1) \), but how well? CRE approach is available but with restrictions on

\[ D(c_i|\{(s_{it}, s_{it}x_{it}) : t = 1, \ldots, T\}). \]
EXAMPLE: Effects of Spending on School Test Pass Rates

. use meap94_98
. tab year

-------------|--------|----------|-------
      yr 1991-2 |        |          |       
  1994     |  1,012 |   14.15  |  14.15
  1995     |  1,205 |   16.85  |  31.01
  1996     |  1,635 |   22.87  |  53.87
  1997     |  1,635 |   22.87  |  76.74
  1998     |  1,663 |   23.26  | 100.00

. egen tobs = sum(1), by(schid)
. tab tobs

number of | Freq.  Percent  Cum.
          |        |          |       
    time  |        |          |       
  periods |        |          |       
  3       |  1,512 |   21.15  |  21.15
  4       |  1,028 |   14.38  |  35.52
  5       |  4,610 |   64.48  | 100.00

. gen tobs3 = tobs == 3
. gen tobs4 = tobs == 4
. egen lavgrexp = mean(lavgrexp), by(schid)
. egen lunchb = mean(lunch), by(schid)
. egen lenrolb = mean(lenrol), by(schid)
. egen y95b = mean(y95), by(schid)
. egen y96b = mean(y96), by(schid)
. egen y97b = mean(y97), by(schid)
. egen y98b = mean(y98), by(schid)
. gen tobs4 = tobs == 4
. gen tobs3 = tobs == 3
. gen tobs3_lavgrexp = tobs3*lavgrexp
. gen tobs4_lavgrexp = tobs4*lavgrexp
. xtreg math4 lavgexp lunch lenrol y95 y96 y97 y98, fe cluster(schid)

Fixed-effects (within) regression
Number of obs = 7150
Group variable: schid
Number of groups = 1683

R-sq: within = 0.3602
between = 0.0292
overall = 0.1514

Obs per group: min = 3
avg = 4.2
max = 5

F(7,1682) = 431.08
Prob > F = 0.0000

(Std. Err. adjusted for 1683 clusters in schid)

|            | Coef.     | Std. Err. | t    | P>|t|   | [95% Conf. Interval] |
|------------|-----------|-----------|------|-------|---------------------|
| math4      | 6.288376  | 2.431317  | 2.59 | 0.010 | 1.519651            |
|            |           |           |      |       | 11.0571             |
| lavgexp    | -0.0215072| 0.0390732 | -0.55| 0.582 | -0.0981445          |
|            |           |           |      |       | 0.05513             |
| lunch      | -2.038461 | 1.789094  | -1.14| 0.255 | -5.547545           |
|            |           |           |      |       | 1.470623            |
| lenrol     | 11.6192   | 0.5358469 | 21.68| 0.000 | 10.56821            |
|            |           |           |      |       | 12.6702             |
| y95        | 13.05561  | 0.6910815 | 18.89| 0.000 | 11.70014            |
|            |           |           |      |       | 14.41108            |
| y96        | 10.14771  | 0.7326314 | 13.85| 0.000 | 8.710745            |
|            |           |           |      |       | 11.58468            |
| y97        | 23.41404  | 0.7669553 | 30.53| 0.000 | 21.90975            |
|            |           |           |      |       | 24.91833            |
| y98        | 11.84422  | 25.16643  | 0.47 | 0.638 | -37.51659           |
|            |           |           |      |       | 61.20503            |

| sigma_u    | 15.84958  |
| sigma_e    | 11.325028 |
| rho        | .66200804 |

(fraction of variance due to u_i)
```
xtreg math4 lavgrea xptobs3_lavgrea xptobs4_lavgrea lunch lenrol
         y95 y96 y97 y98, fe cluster(schid)
```

**Fixed-effects (within) regression**

|                     | Coef.    | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|---------------------|----------|-----------|-------|------|---------------------|
| math4               | 3.501465 | 2.637181  | 1.33  | 0.184| -1.671038 8.673967  |
| lavgrea             | 8.048717 | 4.452136  | 1.81  | 0.071| -0.683593 16.78103 |
| tobs3_lavgrea       | 9.103049 | 5.641861  | 1.61  | 0.107| -1.962758 20.16886 |
| lunch               | -0.292364| 0.387205  | -0.76 | 0.450| -1.051817 0.046709 |
| lenrol              | -2.169307| 1.802788  | -1.20 | 0.229| -5.705251 1.366636 |
| y95                 | 12.01813 | .5312958  | 22.62 | 0.000| 10.97606 13.0602  |
| y96                 | 13.56065 | .6962891  | 19.48 | 0.000| 12.19497 14.92634 |
| y97                 | 10.60934 | .7423073  | 14.29 | 0.000| 9.153396 12.06528 |
| y98                 | 23.84989 | .7789923  | 30.62 | 0.000| 22.322 25.37779  |
| _cons               | 10.6043  | 24.93789  | 0.43  | 0.671| -38.30827 59.51686 |

**F(9,1682) = 337.10, Prob > F = 0.0000**

(Std. Err. adjusted for 1683 clusters in schid)

48
. test tobs3_lavgrexp tobs4_lavgrexp

( 1)  tobs3_lavgrexp = 0
( 2)  tobs4_lavgrexp = 0

F( 2, 1682) = 2.67
  Prob > F = 0.0694
* Now fractional response.

. replace math4 = math4/100
(7150 real changes made)

. replace lunch = lunch/100
(7146 real changes made)

. capture program drop frac_het

. program frac_het
   1.       version 12
   2.       args llf xb zg
   3.       quietly replace `llf' = $ML_y1*log(normal(`xb'*exp(-`zg'))) ///
               + (1 - $ML_y1)*log(1 - normal(`xb'*exp(-`zg')))  
   4.       end

. ml model lf frac_het (math4 = lavgrexp lunch lenrol y95 y96 y97 y98 lavgrexp) ///
   > lunchb lenrolb y95b y96b y97b y98b tobs3 tobs4) (tobs3 tobs4, nocons), ///
   > vce(cluster distid)

. ml max
Number of obs = 7150
Wald chi2(16) = 1690.63
Prob > chi2 = 0.0000

Log pseudolikelihood = -4414.8409

(Std. Err. adjusted for 467 clusters in distid)

| math4 | Coef.  | Std. Err. | z       | P>|z|   | [95% Conf. Interval] |
|-------|--------|-----------|---------|-------|---------------------|
| eq1   |        |           |         |       |                     |
| lavgrexp | 0.1142198 | 0.100819  | 1.13    | 0.257 | -0.0833819 - 0.3118215 |
| lunch  | -0.1396103 | 0.135914  | -1.03   | 0.304 | -0.4059968 - 0.1267763 |
| lenrol | -0.067624  | 0.0736354 | -0.92   | 0.358 | -0.2119468 - 0.0766988 |
| y95    | 0.3241894  | 0.0212756 | 15.24   | 0.000 | 0.2824899 - 0.3658889 |
| y96    | 0.3724917  | 0.0286602 | 13.00   | 0.000 | 0.3163187 - 0.4286647 |
| y97    | 0.2830853  | 0.0301771 | 9.38    | 0.000 | 0.2239392 - 0.3422313 |
| y98    | 0.7162732  | 0.0352366 | 20.33   | 0.000 | 0.6472108 - 0.7853357 |
| lavgrexp | 0.1622914 | 0.0014318 | 12.11   | 0.000 | 0.1000745 - 0.2445745 |
| lunchb | -0.0126246 | 0.0014318 | -8.82   | 0.000 | -0.0154309 - 0.0098183 |
| lenrol | -0.0029272 | 0.0766622 | -0.04   | 0.970 | -0.1531823 - 0.1473279 |
| y95b   | 0.8794288  | 0.76738   | 1.15    | 0.252 | -0.6246084 - 2.383466 |
| y96b   | 0.7270724  | 0.274786  | 3.20    | 0.001 | 0.2812225 - 1.172922 |
| y97b   | 0.6338092  | 0.669099  | 0.95    | 0.342 | -0.6733102 - 1.940929 |
| y98b   | 0.2733774  | 0.639011  | 0.43    | 0.669 | -0.9791201 - 1.525875 |
| tobs3  | 0.022217   | 0.076674  | 0.29    | 0.772 | -0.1280612 - 0.1724953 |
| tobs4  | 0.088465   | 0.122342  | 0.72    | 0.470 | -0.1513217 - 0.3282518 |
| _cons  | -1.856404  | 0.901897 | -2.06   | 0.040 | -3.624091 - 0.0887172 |
| eq2    |        |           |         |       |                     |
| tobs3  | 0.2007713  | 0.0875105 | 2.29    | 0.022 | 0.0292538 - 0.3722888 |
| tobs4  | 0.5504922  | 0.1154134 | 4.77    | 0.000 | 0.3242861 - 0.7766983 |
```
.glm math4 lavgrexp lunch lenrol y95 y96 y97 y98 lavgrexpbi lunchb lenrolb y95b y96b y97b y98b tobs3 tobs4, fam(bin) link(probit) cluster(schid)

note: math4 has noninteger values

Generalized linear models                       No. of obs  =  7150
Optimization        : ML                   Residual df  =  7133

(Std. Err. adjusted for 1683 clusters in schid)

| Robust                      |    Coef. | Std. Err. |   z   |     P>|z| | [95% Conf. Interval] |
|----------------------------|----------|-----------|-------|---------|----------------------|
| math4                      |          |           |       |         |                      |
| lavgrexp                   | 0.1227899| 0.0669842 | 1.83  | 0.067   | -0.0084967 to 0.2540764 |
| lunch                      | -0.0831614| 0.1047519 | -0.79 | 0.427   | -0.2884712 to 0.1221485 |
| lenrol                     | -0.0556512| 0.0490405 | -1.13 | 0.256   | -0.1517689 to 0.0404665 |
| y95                        | 0.3186249| 0.1437888 | 22.16 | 0.000   | 0.2904429 to 0.3468068 |
| y96                        | 0.3647386| 0.1897966 | 19.22 | 0.000   | 0.3275393 to 0.4019379 |
| y97                        | 0.2860664| 0.0201033 | 14.23 | 0.000   | 0.2466647 to 0.3254681 |
| y98                        | 0.6760248| 0.0217182 | 31.13 | 0.000   | 0.6334578 to 0.7185917 |
| lavgrexpbi                 | 0.1658169| 0.08903   | 1.86  | 0.063   | -0.0086787 to 0.3403124 |
| lunchb                     | -0.0113902| 0.010958  | -10.39| 0.000   | -0.0135381 to -0.0092424 |
| lenrolb                    | 0.0202697| 0.0531842 | 0.38  | 0.703   | -0.0839694 to 0.1245088 |
| y95b                       | 0.9325259| 0.3529265 | 2.64  | 0.008   | 0.2408026 to 1.624249 |
| y96b                       | 0.5439736| 0.1438847 | 3.78  | 0.000   | 0.2619647 to 0.8259826 |
| y97b                       | 0.6807815| 0.2587424 | 2.63  | 0.009   | 0.1736557 to 1.187907 |
| y98b                       | 0.2624711| 0.338214  | 0.78  | 0.438   | -0.4004161 to 0.9253584 |
| tobs3                      | -0.0431248| 0.044767 | -0.96 | 0.335   | -0.1308666 to 0.044617 |
| tobs4                      | -0.0771368| 0.0413601 | -1.87 | 0.062   | -0.158201 to 0.0039274 |
| _cons                      | -2.194583| 0.5328879 | -4.12 | 0.000   | -3.239024 to -1.150142 |
```
```
. margins, dydx(lavgrexp)

Average marginal effects                      Number of obs  =  7150
Model VCE      : Robust

Expression : Predicted mean math4, predict()
dy/dx w.r.t. : lavgrexp

------------------------------------------------------------------------------
| Delta-method  | dy/dx      | Std. Err. | z    | P>|z|      | [95% Conf. Interval]    |
------------------------------------------------------------------------------
|               |            |           |     |          |                         |
| lavgrexp      | .043285    | .0236081  | 1.83| 0.067    | -.0029861               |
|               |             |           |     |          | .0895561                |
------------------------------------------------------------------------------
```