Panel Data Methods using Stata
5: Panel - Nonlinear

© A. Colin Cameron
Univ. of Calif. Davis

November 2013

1. Introduction

- Consider nonlinear panel models for short panels.
- One big issue is dealing with individual fixed effects.
- Some models rely more heavily on distributional assumptions than in the linear case.

Outline

2. Cross-section Logit Model Example

- Binary data
  - $y = 1$ if see doctor and $y = 0$ if do not
- Logit model specifies

$$\Pr[y_i = 1|\mathbf{x}_i] = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i'\boldsymbol{\beta})}$$

  - This ensures that $0 < \Pr[\widehat{y_i = 1}|\mathbf{x}_i] < 1$.
- Complication: The marginal effect $\neq \boldsymbol{\beta}$ and varies with $\mathbf{x}$

$$\text{ME}_j = \frac{\partial \Pr[y = 1|\mathbf{x}]}{\partial x_j} = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{(1 + \exp(\mathbf{x}'\boldsymbol{\beta}))^2}\beta_j \neq \beta_j.$$

  - In fact for logit model $\text{ME}_j < 0.25\beta_j$.

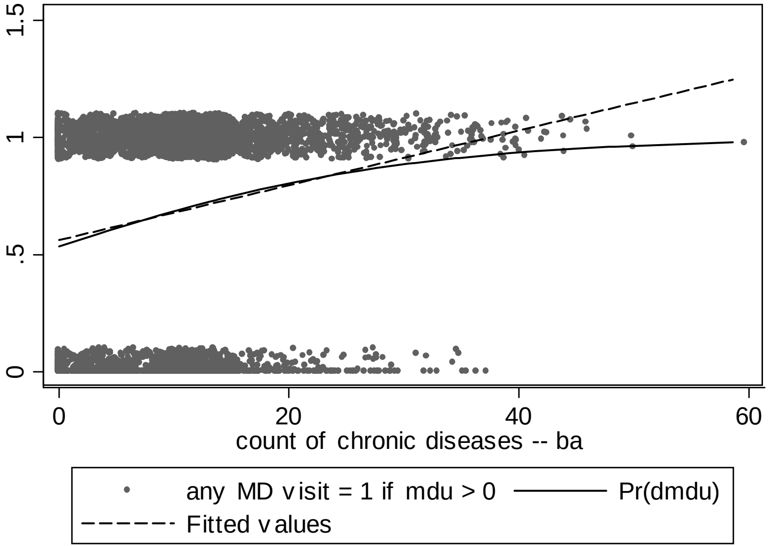
- Rand Health Insurance Experiment data
  - essentially same data as in P. Deb and P.K. Trivedi (2002)
  - $N = 5908$ and individuals may appear for up to five years ($T \leq 5$)
  - use only the fee for service plans (most of sample).
- In this section consider only year one
  - logit model for whether any doctor visit.

```
. use mus18data.dta, clear

.
. keep if year == 1
(14548 observations deleted)

.
. describe dmdu ndisease

              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
dmdu            float   %9.0g                 any MD visit = 1 if mdu > 0
ndisease        float   %9.0g                 count of chronic diseases -- ba

.
. sum dmdu ndisease

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        dmdu |      5638     .693331    .4611517          0          1
    ndisease |      5638    11.20876     6.86571          0       58.6
```

- Logit coefficient is .063 but marginal effect is only 0.013
  - One more chronic disease increases probability of doctor visit by 0.013

```
. logit dmdu ndisease, nolog

Logistic regression                               Number of obs   =       5638
                                                  LR chi2(1)      =     186.27
                                                  Prob > chi2     =     0.0000
Log likelihood = -3382.1798                       Pseudo R2       =     0.0268

------------------------------------------------------------------------------
        dmdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    ndisease |   .0634046    .004939    12.84   0.000     .0537243    .0730849
       _cons |   .1423681   .0577308     2.47   0.014     .0292177    .2555185
------------------------------------------------------------------------------

.
. margins, dydx(*)

Average marginal effects                          Number of obs   =       5638
Model VCE    : OIM

Expression   : Pr(dmdu), predict()
dy/dx w.r.t. : ndisease

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    ndisease |   .0130581   .0009696    13.47   0.000     .0111576    .0149585
------------------------------------------------------------------------------
```

- Plot of data (jittered), logit prediction and OLS prediction.



3. Nonlinear Cross-section Models Summary

- Consider three common nonlinear models.
- 1. Binary data
  - $y = 0$ or 1 such as employed $(y = 1)$ or not
  - logit model specifies $\Pr[y_i = 1|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})/[1 + \exp(\mathbf{x}_i'\boldsymbol{\beta})]$
  - probit model specifies $\Pr[y_i = 1|\mathbf{x}_i] = \Phi(\mathbf{x}_i'\boldsymbol{\beta})$ where $\Phi(\cdot)$ is the standard normal c.d.f.
- 2. Count data
  - $y = 0, 1, 2, \ldots$ such as number of doctor visits
  - Poisson model with conditional mean $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$
- 3. Tobit model
  - $y_i^* \sim N[\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2]$
  - but we only observe $y_i = y_i^*$ if $y_i^* > 0$ and $y_i = 0$ if $y_i^* \leq 0$
- First two are special cases of generalized linear models
- Third is very dependent on distributional assumptions.

4. Panel Data Example: Health Care Utilization

- Rand Health Insurance Experiment data
  - $N = 5908$ and $T = 5$.
- Consider three different types of dependent variable
  - `dmdu` is binary; `med` is continuous ≥ 0; `mdu` is count.

```
. describe dmdu med mdu lcoins ndisease female age lfam child id year

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------------
dmdu            float  %9.0g                  any MD visit = 1 if mdu > 0
med             float  %9.0g                  medical exp excl outpatient men
mdu             float  %9.0g                  number face-to-fact md visits
lcoins          float  %9.0g                  log(coinsurance+1)
ndisease        float  %9.0g                  count of chronic diseases -- ba
female          float  %9.0g                  female
age             float  %9.0g                  age that year
lfam            float  %9.0g                  log of family size
child           float  %9.0g                  child
id              float  %9.0g                  person id, leading digit is sit
year            float  %9.0g                  study year
```

```
. * Summarize dependent variables and regressors
. summarize dmdu med mdu lcoins ndisease female age lfam child id year
```

Variable	Obs	Mean	Std. Dev.	Min	Max
dmdu	20186	.6875062	.4635214	0	1
med	20186	171.5892	698.2689	0	39182.02
mdu	20186	2.860696	4.504765	0	77
lcoins	20186	2.383588	2.041713	0	4.564348
ndisease	20186	11.2445	6.741647	0	58.6
female	20186	.5169424	.4997252	0	1
age	20186	25.71844	16.76759	0	64.27515
lfam	20186	1.248404	.5390681	0	2.639057
child	20186	.4014168	.4901972	0	1
id	20186	357971.2	180885.6	125024	632167
year	20186	2.420044	1.217237	1	5

- Most are in for first 3 years or all 5 years.

```
. xtdescribe

      id:  125024, 125025, ..., 632167                       n =       5908
    year:  1, 2, ..., 5                                      T =          5
           Delta(year) = 1 unit
           Span(year)  = 5 periods
           (id*year uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       2       3         3         5       5       5

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------
     3710     62.80   62.80 |  111..
     1584     26.81   89.61 |  11111
      156      2.64   92.25 |  1....
      147      2.49   94.74 |  11...
       79      1.34   96.07 |  ..1..
       66      1.12   97.19 |  .11..
       33      0.56   97.75 |  ..111
       33      0.56   98.31 |  .1111
       29      0.49   98.80 |  ...11
       71      1.20  100.00 | (other patterns)
 ---------------------------+---------
     5908    100.00         |  XXXXX
```

- Very little within variation here
  - so fixed effects estimator just illustrative
  - plus little need anyway as designed experiment
  - to the extent that the experiment changes regressors over time this will not be well picked up on by a time-invariant fixed effect.

```
. * Panel summary of time-varying regressors
. xtsum age lfam child

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
age      overall |  25.71844   16.76759          0   64.27515 |     N =   20186
         between |             16.97265          0   63.27515 |     n =    5908
         within  |             1.086687   23.46844   27.96844 | T-bar = 3.41672
                 |                                            |
lfam     overall |  1.248404   .5390681          0   2.639057 |     N =   20186
         between |             .5372082          0   2.639057 |     n =    5908
         within  |             .0730824   .3242075    2.44291 | T-bar = 3.41672
                 |                                            |
child    overall |  .4014168   .4901972          0          1 |     N =   20186
         between |             .4820984          0          1 |     n =    5908
         within  |             .1096116  -.3985832   1.201417 | T-bar = 3.41672
```

5. Panel Logit

- 68% visited doctor at least once in year.
- There is reasonable within variation.
    - Also autocorrelation is .38 at one lag and .36 at two lags.

```
. * Panel summary of dependent variable
. xtsum dmdu

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
dmdu     overall |  .6875062    .4635214         0          1 |     N =   20186
         between |              .3571059         0          1 |     n =    5908
         within  |              .3073307  -.1124938   1.487506 | T-bar = 3.41672

.
. * Year-to-year transitions in whether visit doctor
. xttrans dmdu

      any MD |  any MD visit = 1 if
   visit = 1 |        mdu > 0
  if mdu > 0 |        0          1 |     Total
-------------+----------------------+----------
           0 |    58.87      41.13 |    100.00
           1 |    19.73      80.27 |    100.00
-------------+----------------------+----------
       Total |    31.81      68.19 |    100.00
```

Pooled Logit

- Assume that usual binary logit model still applies with panel data

$$\Pr[y_{it} = 1 | \mathbf{x}_{it}] = \frac{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_{it}'\boldsymbol{\beta})}$$

- But now allow y_{it} to be correlated over time for given i
  - $\text{Cor}[y_{it}, y_{is}] \neq 0$ though $\text{Cor}[y_{it}, y_{js}] = 0$ for $j \neq i$.
- Then usual logit MLE is still consistent
  - but need to use cluster-robust standard errors (assume $N \rightarrow \infty$).

- Cluster-robust standard errors are 40-50% larger than default

```
. logit dmdu lcoins ndisease female age lfam child, vce(cluster id) nolog

Logistic regression                               Number of obs   =      20186
                                                  Wald chi2(6)    =     488.18
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -11973.392                 Pseudo R2       =     0.0450

                               (Std. Err. adjusted for 5908 clusters in id)
------------------------------------------------------------------------------
             |               Robust
        dmdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.1572107   .0109064   -14.41   0.000    -.1785869   -.1358345
    ndisease |    .050301   .0039657    12.68   0.000     .0425285    .0580735
      female |   .3091573   .0445772     6.94   0.000     .2217876     .396527
         age |   .0042689   .0022307     1.91   0.056    -.0001032     .008641
        lfam |  -.2047573   .0470287    -4.35   0.000    -.2969318   -.1125828
       child |   .0921709   .0728107     1.27   0.206    -.0505355    .2348773
       _cons |   .6039409   .1107712     5.45   0.000     .3868333    .8210485
------------------------------------------------------------------------------
```

Pooled Logit with Exchangeable Errors

- Again assume that usual binary logit model still applies with panel data

$$\Pr[y_{it} = 1|\mathbf{x}_{it}] = \frac{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_{it}'\boldsymbol{\beta})}$$

- But now specify a model for the correlation of y_{it}
  - $\text{Cor}[y_{it}, y_{is}] = \rho$ for $s \neq t$
  - Again $\text{Cor}[y_{it}, y_{js}] = 0$ for $j \neq i$.
- Advantage is more efficient estimation possibly
  - but again use cluster-robust standard errors (assume $N \rightarrow \infty$) to guard against wrong error correlation model.

- Similar coefficients, slightly smaller standard errors.

```
. xtlogit dmdu lcoins ndisease female age lfam child, pa corr(exch) vce(robust)

GEE population-averaged model                   Number of obs      =     20186
Group variable:                         id      Number of groups   =      5908
Link:                                logit      Obs per group: min =         1
Family:                           binomial                     avg =       3.4
Correlation:                  exchangeable                     max =         5
                                                Wald chi2(6)       =    521.45
Scale parameter:                         1      Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
        dmdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.1603179   .0107779   -14.87   0.000    -.1814422   -.1391935
    ndisease |   .0515445   .0038528    13.38   0.000     .0439931    .0590958
      female |   .2977003   .0438316     6.79   0.000      .211792    .3836086
         age |   .0045675   .0021001     2.17   0.030     .0004514    .0086836
        lfam |  -.2044045   .0455004    -4.49   0.000    -.2935837   -.1152254
       child |   .1184697   .0674367     1.76   0.079    -.0137039    .2506432
       _cons |   .5776986    .106591     5.42   0.000      .368784    .7866132
------------------------------------------------------------------------------
```

Random Effects Logit

- Now assume a different model:

$$\begin{aligned} \Pr[y_{it} &= 1|\alpha_i, \mathbf{x}_{it}] = \frac{\exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})}{1 + \exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})} \\ \alpha_i &\sim N[0, \sigma_\alpha^2] \end{aligned}$$

- Then no longer get the logit probability

$$\begin{aligned} \Pr[y_{it} &= 1|\mathbf{x}_{it}] = \int \frac{\exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})}{1 + \exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})} f(a_i) d\alpha_i \\ &\neq \frac{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_{it}'\boldsymbol{\beta})} \end{aligned}$$

- And need to use numerical methods as no closed form solution for log-density
  - but easy numerically as one-dimensional integral
  - use Gaussian quadrature

- Coefficients about 50% larger than using logit

```
. xtlogit dmdu lcoins ndisease female age lfam child, re nolog

Random-effects logistic regression              Number of obs      =     20186
Group variable: id                              Number of groups   =      5908

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       3.4
                                                               max =         5

Integration method: mvaghermite                 Integration points =        12

                                                Wald chi2(6)       =    549.76
Log likelihood  = -10878.687                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        dmdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.2403864   .0162836   -14.76   0.000    -.2723017    -.208471
    ndisease |    .078151   .0055456    14.09   0.000     .0672819    .0890201
      female |   .4631005   .0663209     6.98   0.000     .3331138    .5930871
         age |   .0073441   .0031508     2.33   0.020     .0011687    .0135194
        lfam |  -.3021841   .0644721    -4.69   0.000    -.4285471    -.175821
       child |   .1935357   .1002267     1.93   0.053     -.002905    .3899763
       _cons |   .8629898   .1568968     5.50   0.000     .5554778    1.170502
-------------+----------------------------------------------------------------
    /lnsig2u |   1.225652   .0490898                      1.129438    1.321866
-------------+----------------------------------------------------------------
     sigma_u |    1.84564    .045301                      1.758953    1.936599
         rho |   .5087003   .0122687                      .4846525     .532708
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  2189.41 Prob >= chibar2 = 0.000
```

Fixed Effects Logit

- Now assume like random effects

$$\Pr[y_{it} = 1|\alpha_i, \mathbf{x}_{it}] = \frac{\exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})}{1 + \exp(\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})}$$

- But now allow $\mathrm{Cor}[\alpha_i, \mathbf{x}_{it}] = 0$.
  - Hard to eliminate α_i but it is possible.
- Can do MLE conditional on $\sum_t y_{it}$ (for $\sum_t y_{it} \neq 0$ and $\sum_t y_{it} \neq T$)
  - requires individual to have at least one move from 0 to 1 or 1 to 0.
  - this works for logit but not for probit.

- To see how α_i is eliminated consider the following.
- Suppose $T = 2$ and $\sum_t y_{it} = 1$
  - then either sequence 01 or 10 is possible and α_i drops out

$$
\begin{aligned}
&\Pr[y_{i1}=0, y_{i2}=1 | y_{i1}+y_{i2}=1] \\
&= \Pr[y_{i1}=0, y_{i2}=1] / \{\Pr[y_{i1}=0, y_{i2}=1] + \Pr[y_{i1}=1, y_{i2}=0]\} \\
&= \frac{\{e^{\alpha_i+\mathbf{x}_{i1}'\beta}/(1+e^{\alpha_i+\mathbf{x}_{i1}'\beta})\}\times\{1/(1+e^{\alpha_i+\mathbf{x}_{i2}'\beta})\}}{\left[\{e^{\alpha_i+\mathbf{x}_{i1}'\beta}/(1+e^{\alpha_i+\mathbf{x}_{i1}'\beta})\}\times\{1/(1+e^{\alpha_i+\mathbf{x}_{i2}'\beta})\}\right]\times\left[\{1/(1+e^{\alpha_i+\mathbf{x}_{i1}'\beta})\}\times 1+\{e^{\alpha_i+\mathbf{x}_{i2}'\beta}/(\right.} \\
&= \frac{\exp(\mathbf{x}_{i1}'\beta)}{\exp(\mathbf{x}_{i1}'\beta)+\exp(\mathbf{x}_{i2}'\beta)} \\
&= \frac{\exp((\mathbf{x}_{i1}-\mathbf{x}_{i2})'\beta)}{1+\exp((\mathbf{x}_{i1}-\mathbf{x}_{i2})'\beta)}.
\end{aligned}
$$

  - So do MLE conditional on $\sum_t y_{it} = 1$ as logit model with regressor $\mathbf{x}_{i1} - \mathbf{x}_{i2}$.

- Note that time-invariant regressors are dropped, as expected

```
. * Logit fixed-effects estimator
. xtlogit dmdu lcoins ndisease female age lfam child, fe nolog
note: multiple positive outcomes within groups encountered.
note: 3459 groups (11161 obs) dropped because of all positive or
      all negative outcomes.
note: lcoins omitted because of no within-group variance.
note: ndisease omitted because of no within-group variance.
note: female omitted because of no within-group variance.
```

```
Conditional fixed-effects logistic regression   Number of obs      =      9025
Group variable: id                              Number of groups   =      2449

                                                Obs per group: min =         2
                                                               avg =       3.7
                                                               max =         5

                                                LR chi2(3)         =     10.74
Log likelihood  = -3395.5996                    Prob > chi2        =    0.0132

------------------------------------------------------------------------------
        dmdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  (omitted)
    ndisease |  (omitted)
      female |  (omitted)
         age |  -.0341815   .0183827    -1.86   0.063     -.070211     .001848
        lfam |    .478755   .2597327     1.84   0.065    -.0303116    .9878217
       child |    .270458   .1684974     1.61   0.108    -.0597907    .6007068
------------------------------------------------------------------------------
```

Estimator Comparison

```
. * Logit mixed-effects estimator (same as xtlogit, re)
. * xtmelogit dmdu lcoins ndisease female age lfam child || id:
.
. * Panel logit estimator comparison
. global xlist lcoins ndisease female age lfam child

. quietly logit dmdu $xlist, vce(cluster id)

. estimates store POOLED

. quietly xtlogit dmdu $xlist, pa corr(exch) vce(robust)

. estimates store PA

. quietly xtlogit dmdu $xlist, re     // SEs are not cluster-robust

. estimates store RE

. quietly xtlogit dmdu $xlist, fe     // SEs are not cluster-robust

. estimates store FE

. estimates table POOLED PA RE FE, equations(1) se b(%8.4f) stats(N ll) stfmt(%8.0f)
```

Variable	POOLED	PA	RE	FE
#1				
lcoins	-0.1572	-0.1603	-0.2404	(omitted)
	0.0109	0.0108	0.0163	
ndisease	0.0503	0.0515	0.0782	(omitted)
	0.0040	0.0039	0.0055	
female	0.3092	0.2977	0.4631	(omitted)
	0.0446	0.0438	0.0663	
age	0.0043	0.0046	0.0073	-0.0342
	0.0022	0.0021	0.0032	0.0184
lfam	-0.2048	-0.2044	-0.3022	0.4788
	0.0470	0.0455	0.0645	0.2597
child	0.0922	0.1185	0.1935	0.2705
	0.0728	0.0674	0.1002	0.1685
_cons	0.6039	0.5777	0.8630	
	0.1108	0.1066	0.1569	
lnsig2u				
_cons			1.2257	
			0.0491	
Statistics				
N	20186	20186	20186	9025
ll	-11973		-10879	-3396

legend: b/se

- Individual fixed effects cannot be eliminated from probit $\Pr[y_{it} = 1|\alpha_i, \mathbf{x}_{it}, \boldsymbol{\beta}] = \Phi(\alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta})$ (unlike logit)
- Instead can do dummy variables probit
  - coefficient estimates are biased
  - but Fernández-Val (2009) shows that there is only small bias in ratio of coefficients $\widehat{\beta}_j/\widehat{\beta}_k$ and in computed average marginal effects even for small T
  - here $\widehat{\text{AME}} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{i=1}^{T}\phi(\widehat{\alpha}_i + \mathbf{x}'_{it}\widehat{\boldsymbol{\beta}})$
- For individual random effects that are normally distributed can show
  - $\Pr[y_{it} = 1|\alpha_i, \mathbf{x}_{it}, \boldsymbol{\beta}] = \Phi(\alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta})$ where $\alpha_i \sim N[0, \sigma^2_\alpha]$
  - $\Longrightarrow \Pr[y_{it} = 1|\mathbf{x}_{it}, \boldsymbol{\beta}] = \int \Pr[y_{it} = 1|\alpha_i, \mathbf{x}_{it}, \boldsymbol{\beta}]f(\alpha_i)d\alpha_i = \Phi(\mathbf{x}'_{it}\boldsymbol{\beta}/\sqrt{1+\sigma^2_\alpha})$ so same as regular probit with coefficients scaled by $\sqrt{1+\sigma^2_\alpha}$.

3. Nonlinear Panel Models Overview

- General approaches similar to linear case
  - Pooled estimation or population-averaged
  - Random effects
  - Fixed effects
- Complications
  - Random effects often not tractable so need numerical integration
  - Fixed effects models in short panels are generally not estimable due to the incidental parameters problem.
- Here we consider short panels throughout.
- Standard nonlinear models are:
  - Binary: logit and probit
  - Counts: Poisson and negative binomial
  - Truncated: Tobit

Nonlinear Panel Models

- 1. A pooled or population-averaged model may be used.
  - This is same model as in cross-section case, with adjustment for correlation over time for a given individual.
- 2. A fully parametric model with individual additive effect α_i has conditional density

$$f(y_{it}|\alpha_i, \mathbf{x}_{it}) = f(y_{it}, \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}, \gamma), \quad t = 1, ..., T_i, \; i = 1, ..., N,$$

  where γ denotes additional model parameters such as variance parameters and α_i is an individual effect.
- 3. A conditional mean model with individual effects

$$\mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}] = \alpha_i + g(\mathbf{x}_{it}'\boldsymbol{\beta})$$

  For models with $0 < \mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}] < \infty$ can also have multiplicative effects

$$\mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}] = \alpha_i \times g(\mathbf{x}_{it}'\boldsymbol{\beta}).$$

Marginal Effects: Cross-section Model

- Start with a nonlinear cross-section model with $\mathrm{E}[y_i|\mathbf{x}_i] = g(\mathbf{x}_i'\boldsymbol{\beta})$.
- We are often interested in the marginal effects

$$\frac{\partial \mathrm{E}[y|\mathbf{x}]}{\partial \mathbf{x}} = \frac{\partial g(\mathbf{x}'\boldsymbol{\beta})}{\partial \mathbf{x}} \neq \boldsymbol{\beta}.$$

- This is no longer just $\boldsymbol{\beta}$ and it varies with evaluation point $\mathbf{x}$.
  - AME Average marginal effect: $\frac{1}{N}\sum_{i=1}^{N} \frac{\partial g(\mathbf{x}_i'\boldsymbol{\beta})}{\partial \mathbf{x}_i}$.
  - MEM Marginal effect at mean $\bar{\mathbf{x}}$: $\left.\frac{\partial g(\mathbf{x}'\boldsymbol{\beta})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\bar{\mathbf{x}}}$.
  - MER Marginal effect at representative value $\mathbf{x}^*$: $\left.\frac{\partial g(\mathbf{x}'\boldsymbol{\beta})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^*}$.
- Single-index model $\mathrm{E}[y|\mathbf{x}] = g(\mathbf{x}'\boldsymbol{\beta})$ is specialization of $g(\mathbf{x}, \boldsymbol{\beta})$
  - Ratio of coefficient equals the ratio of marginal effects
  - $\frac{\partial \mathrm{E}[y|\mathbf{x}]/\partial x_j}{\partial \mathrm{E}[y|\mathbf{x}]/\partial \mathbf{x}_k} = \frac{g'(\mathbf{x}'\boldsymbol{\beta})\beta_j}{g'(\mathbf{x}'\boldsymbol{\beta})\beta_k} = \frac{\beta_j}{\beta_k}$.

Marginal Effects: Panel Model

- For pooled panel model $\mathrm{E}[y_{it}|\mathbf{x}_{it}] = g(\mathbf{x}_{it}'\boldsymbol{\beta})$ the same issues arise.
- For individual effects panel models the marginal effect can also depend on α_i!
- e.g. For multiplicative individual effects model

$$\mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}] = \alpha_i \times g(\mathbf{x}_{it}'\boldsymbol{\beta})$$

$$\Longrightarrow \quad \partial\mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}]/\partial\mathbf{x}_{it} = \alpha_i \times \partial g(\mathbf{x}_{it}'\boldsymbol{\beta})/\partial\mathbf{x}_{it}$$

- Even if we can consistently estimate $\boldsymbol{\beta}$ in the presence of α_i, we can't get the marginal effect
  - but might get the ratio of ME's
  - see also discussion for AME's in panel binary probit.

7. Pooled or Population-averaged estimation

- Extend pooled OLS
  - Give the usual cross-section command for conditional mean models or conditional density models but then get cluster-robust standard errors
  - Probit example:

    ```
    probit y x, vce(cluster id)
    ```
    or
    ```
    xtgee y x, fam(binomial) link(probit) corr(ind)
    vce(cluster id)
    ```
- Extend pooled feasible GLS
  - Estimate with an assumed correlation structure over time
  - Equicorrelated probit example:

    ```
    xtprobit y x, pa vce(boot)
    ```
    or
    ```
    xtgee y x, fam(binomial) link(probit) corr(exch)
    vce(cluster id)
    ```

Random Effects Estimation

- Assume individual-specific effect α_i has specified distribution $g(\alpha_i|\boldsymbol{\eta})$.
- Then the unconditional density for the i^{th} observation is

$$
\begin{aligned}
& f(y_{it}, ..., y_{iT}|\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}, \boldsymbol{\beta}, \gamma, \boldsymbol{\eta}) \\
= & \int \left[\prod\nolimits_{t=1}^{T} f(y_{it}|\mathbf{x}_{it}, \alpha_i, \boldsymbol{\beta}, \gamma)\right] g(\alpha_i|\boldsymbol{\eta}) d\alpha_i
\end{aligned}
$$

- Analytical solution:
  - For Poisson with gamma random effect
  - For negative binomial with gamma effect
  - Use `xtpoisson, re` and `xtnbreg, re`
- No analytical solution:
  - For other models.
  - Instead use numerical integration (only univariate integration is required).
  - Assume normally distributed random effects.
  - Use `re` option for `xtlogit`, `xtprobit`
  - Use `normal` option for `xtpoisson` and `xtnbreg`

Random Slopes Estimation

- Can extend to random slopes.
  - Nonlinear generalization of `xtmixed`
  - Then higher-dimensional numerical integral.
  - Use adaptive Gaussian quadrature
- Stata commands are:
  - `xtmelogit` for binary data
  - `xtmepoisson` for counts
- Stata add-on that is very rich:
  - `gllamm` (generalized linear and latent mixed models)
  - Developed by Sophia Rabe-Hesketh and Anders Skrondal.

Fixed Effects Estimation

- In general not possible in short panels.
- Incidental parameters problem:
  - N fixed effects α_i plus K regressors means $(N+K)$ parameters
  - But $(N+K) \rightarrow \infty$ as $N \rightarrow \infty$
  - Need to eliminate α_i by some sort of differencing
  - possible for Poisson, negative binomial and logit
  - also some bias-corrected estimators have been suggested
- Stata commands
  - `xtlogit, fe`
  - `xtpoisson, fe vce(robust)`
  - `xtnbreg, fe`
- Fixed effects extended to dynamic models for logit and probit.
  - No Stata command.

Correlated Random Effects Estimation

- Suppose time-invariant individual effect depends on the average $\bar{\mathbf{x}}_i$
  - $\alpha_i = \pi_0 + \bar{\mathbf{x}}_i'\boldsymbol{\pi} + v_i; \; v_i \sim N[0, \sigma_v^2]$
- The usual random effects model is
  - density $f(y_{it}|\alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})$ where $\alpha_i \sim N[0, \sigma_\alpha^2]$
- We instead estimate
  - density $f(y_{it}|v_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \bar{\mathbf{x}}_i'\boldsymbol{\pi})$ where $v_i \sim N[0, \sigma_v^2]$
  - so do the usual nonlinear RE estimator with $\bar{\mathbf{x}}_i$ added as a regressor
- Notes
  - can only identify coefficients of time-varying regressors
  - can generalize from $\bar{\mathbf{x}}_i$ to other exchangeable functions of data (Altonji and Matzkin (2005)
  - Wooldridge considers extension to unbalanced panel data

8. Panel Count

- Model annual number of doctor visits

```
. * Panel summary of dependent variable
. xtsum mdu

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
mdu      overall |  2.860696   4.504765          0         77 |     N =   20186
         between |             3.785971          0   63.33333 |     n =    5908
         within  |             2.575881  -34.47264    40.0607 | T-bar = 3.41672
```

- Data are counts
  - Overdispersed as variance $= (4.51)^2 = 20.34$ is 7 times the mean.
- For conditional mean modelling Poisson may still be okay.
- For parametric modelling need richer model e.g. negative binomial.
- Cross section case Poisson specifies $\mathrm{E}[y_i|\mathbf{x}_i'\boldsymbol{\beta}] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$
  - Poisson MLE is quite robust as first-order conditions are

$$\sum_{i=1}^{N}(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i = \mathbf{0}.$$

  - But be sure to use robust standard errors.

- Poisson model with individual specific effect

$$\begin{aligned} \mathrm{E}\left[y_{it}|\delta_i, \mathbf{x}_{i1}, ..., \mathbf{x}_{i,T}\right] &= \exp(\delta_i + \mathbf{x}_{it}'\boldsymbol{\beta}) \\ &= \exp(\delta_i)\exp(\mathbf{x}_{it}'\boldsymbol{\beta}) \\ &= \alpha_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta}). \end{aligned}$$

- This is a multiplicative fixed effect that can be quasi-differenced out.
- Averaging over t for given i

$$\mathrm{E}\left[\bar{y}_i|\delta_i, \mathbf{x}_{i1}, ..., \mathbf{x}_{i,T}\right] = \alpha_i \overline{\exp(\mathbf{x}_i'\boldsymbol{\beta})}$$

- Quasi-difference of the two equations

$$\begin{aligned} \mathrm{E}\left[y_{it} - \frac{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}{\overline{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}} \times \bar{y}_i | \mathbf{x}_{i1}, ..., \mathbf{x}_{iT}\right] &= 0 \\ \mathrm{E}\left[\mathbf{x}_{it}\left(y_{it} - \frac{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}{\overline{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}} \times \bar{y}_i\right)\right] &= 0.. \end{aligned}$$

- So m-estimator for $\boldsymbol{\beta}$ solves

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\left(y_{it}-\frac{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}{\overline{\exp(\mathbf{x}_{it}'\boldsymbol{\beta})}}\times\bar{y}_i\right)=0.$$

- Same as fixed estimator in the fully parametric Poisson case!
  - But be sure to use robust standard errors.

```
. xtpoisson mdu lcoins ndisease female age lfam child, fe vce(robust)
note: 265 groups (265 obs) dropped because of only one obs per group
note: 666 groups (2130 obs) dropped because of all zero outcomes
note: lcoins dropped because it is constant within group
note: ndisease dropped because it is constant within group
note: female dropped because it is constant within group

Iteration 0:   log pseudolikelihood = -24182.852
Iteration 1:   log pseudolikelihood = -24173.211
Iteration 2:   log pseudolikelihood = -24173.211

Conditional fixed-effects Poisson regression    Number of obs      =     17791
Group variable: id                              Number of groups   =      4977

                                                Obs per group: min =         2
                                                               avg =       3.6
                                                               max =         5

                                                Wald chi2(3)       =      4.58
Log pseudolikelihood  = -24173.211              Prob > chi2        =    0.2051

                                   (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0112009   .0091493    -1.22   0.221    -.0291331    .0067314
        lfam |   .0877134   .1160837     0.76   0.450    -.1398064    .3152332
       child |   .1059867   .0786326     1.35   0.178    -.0481304    .2601037
------------------------------------------------------------------------------
```

Poisson Estimator Comparison

```
. * Comparison of Poisson panel estimators
. quietly xtpoisson mdu lcoins ndisease female age lfam child, pa corr(unstr) vce(robust)

. estimates store PPA_ROB

. quietly xtpoisson mdu lcoins ndisease female age lfam child, re

. estimates store PRE

. quietly xtpoisson mdu lcoins ndisease female age lfam child, re normal

. estimates store PRE_NORM

. quietly xtpoisson mdu lcoins ndisease female age lfam child, fe vce(robust)

. estimates store PFE_ROB

. quietly xtpoisson mdu lcoins ndisease female age lfam child, fe

. estimates store PFE

. estimates table PPA_ROB PRE PRE_NORM PFE_ROB PFE, equations(1) b(%8.4f) se stats(N ll) s
> tfmt(%8.0f)
```

Variable	PPA_ROB	PRE	PRE_NORM	PFE_ROB	PFE
#1					
lcoins	-0.0804	-0.0878	-0.1145		
	0.0078	0.0069	0.0073		
ndisease	0.0346	0.0388	0.0409		
	0.0024	0.0022	0.0023		
female	0.1585	0.1667	0.2084		
	0.0334	0.0286	0.0305		
age	0.0031	0.0019	0.0027	-0.0112	-0.0112
	0.0015	0.0011	0.0012	0.0091	0.0039
lfam	-0.1407	-0.1352	-0.1443	0.0877	0.0877
	0.0294	0.0260	0.0265	0.1161	0.0555
child	0.1014	0.1083	0.0737	0.1060	0.1060
	0.0430	0.0341	0.0345	0.0786	0.0438
_cons	0.7765	0.7574	0.2873		
	0.0717	0.0618	0.0642		
lnalpha					
_cons		0.0251			
		0.0210			
lnsig2u					
_cons			0.0550		
			0.0255		
Statistics					
N	20186	20186	20186	17791	17791
ll		-43241	-43227	-24173	-24173

legend: b/se

9. Panel Tobit

- Model annual medical expenditure

```
. * Panel summary of dependent variable
. xtsum med

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
med      overall |  171.5892   698.2689          0   39182.02 |     N =   20186
         between |             503.2589          0   19615.14 |     n =    5908
         within  |              526.269  -19395.28    20347.2 | T-bar = 3.41672
```

- Not shown is that many zeroes (22% of observations)
  - leads to inconsistent estimator
- Control for selection assuming highly parametric Tobit model

$$
\begin{aligned}
y_{it}^{*} &= \alpha_i + \mathbf{x}_{it}'\beta + \varepsilon_{it} \\
y_{it} &= \begin{cases} y_{it}^{*} & y_{it}^{*} > 0 \\ 0 & y_{it}^{*} \leq 0 \end{cases} \\
\varepsilon_{it} &\sim \mathcal{N}[0, \sigma_{\varepsilon}^{2}] \text{ and } \alpha_i \sim \mathcal{N}[0, \sigma_{\alpha}^{2}].
\end{aligned}
$$

```
. * Tobit random-effects estimator
. xttobit med lcoins ndisease female age lfam child, ll(0) nolog

Random-effects tobit regression                 Number of obs      =     20186
Group variable: id                              Number of groups   =      5908

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       3.4
                                                               max =         5

                                                Wald chi2(6)       =    573.45
Log likelihood  = -130030.45                    Prob > chi2        =    0.0000
```

med	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lcoins	-31.10247	3.578498	-8.69	0.000	-38.1162	-24.08875
ndisease	13.49452	1.139156	11.85	0.000	11.26182	15.72722
female	60.10112	14.95966	4.02	0.000	30.78072	89.42152
age	4.075582	.7238253	5.63	0.000	2.656911	5.494254
lfam	-57.75023	14.68422	-3.93	0.000	-86.53077	-28.96968
child	-52.02314	24.21619	-2.15	0.032	-99.48599	-4.560284
_cons	-98.27203	36.05977	-2.73	0.006	-168.9479	-27.59618
/sigma_u	371.3134	8.64634	42.94	0.000	354.3668	388.2599
/sigma_e	715.1779	4.704581	152.02	0.000	705.9571	724.3987
rho	.2123246	.0086583			.1957541	.2296872

```
  Observation summary:         4453  left-censored observations
                              15733     uncensored observations
                                  0 right-censored observations
```

10. Summary

- Stata commands cover the main models

	Counts	**Binary**
Pooled	poisson	logit
	nbreg	probit
GEE (PA)	xtgee,family(poisson)	xtgee,family(binomial) link(logit
	xtgee,family(nbinomial)	xtgee,family(poisson) link(probit
RE	xtpoisson, re	xtlogit, re
	xtnbreg, fe	xtprobit, re
Random slopes	xtmepoisson	xtmelogit
FE	xtpoisson, fe	xtlogit, fe
	xtnbreg, fe	

plus `tobit` and `xttobit`.

11. References

- Fernández-Val, I. (2009), "Fixed effects estimation of structural parameters and marginal effects in panel probit models," Journal of Econometrics, 71-85.