# Thursday
# Counts: Panel Data

A. Colin Cameron
Univ. of Calif. - Davis

. .

Nonlinear Cross-section and Panel Regression
Models for Count Data

. .

Queens University, Canada
Department of Economics

May 13-16, 2019

# 1. Introduction

- Count data models are for dependent variable $y = 0, 1, 2, \ldots$
- Previously considered cross-section models.
- Now consider panel data
  - focus on short panels with large $N$ and small $T$.
  - first static models and then dynamic models.

# Outline

# 2. Panel Data Summary: Reading in data

- Data organization may be
  - long form: each observation is an individual-time $(i, t)$ pair
  - wide form: each observation is data on $i$ for all time periods
  - wide form: each observation is data on $t$ for all individuals
- `xt` commands require data in long form
  - use `reshape long` command to convert from wide to long form
  - see Cameron and Trivedi (2010) chapter 8.11.
- Data here are already in long form
  - `use mus18data.dta, clear`

# Data example: Doctor visits (RAND)

- Data from RAND health insurance experiment.
  - $y$ is number of doctor visits.

```
. use mus18data.dta, clear

. describe mdu lcoins ndisease female age lfam child id year

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------
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 Data using Non-panel Commands

- Dependent variable mdu is very overdispersed: $\widehat{V}[y] = 4.50^2 \simeq 7 \times \bar{y}$.

```
. summarize mdu lcoins ndisease female age lfam child id year
```

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| 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 |

# Stata Commands for Panel Data Summary

- Commands `describe`, `summarize` and `tabulate` confound cross-section and time series variation.
- Instead use specialized panel commands after `xtset`:
  - `xtdescribe`: extent to which panel is unbalanced
  - `xtsum`: separate within (over time) and between (over individuals) variation
  - `xttab`: tabulations within and between for discrete data e.g. binary
  - `xttrans`: transition frequencies for discrete data
  - `xtline`: time series plot for each individual on one chart
  - `xtdata`: scatterplots for within and between variation.

- Declare as panel data: `xtset id year` and then `xtdescribe`
  - most in for 3 years or for 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
```

# Within and between variation

- Decompose variation
  - between is across individuals $s^2_{between} = \frac{1}{N}\sum_i (\bar{y}_i - \bar{\bar{y}})^2$
    - this is zero for `year` which is the same for each individual
  - within is over time for a given individual $s^2_{within} = \frac{1}{N}\sum_i (y_{it} - \bar{y}_i)^2$
    - this is zero for a variable that does not vary over time
  - note
    $s^2_{between} + s^2_{within} = 3.786^2 + 12.576^2 = 4.579^2 \simeq s^2_{overall} = 4.504^2$.

```
. * 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
```

- `xttrans` provides transition probabilities for discrete-valued variable.

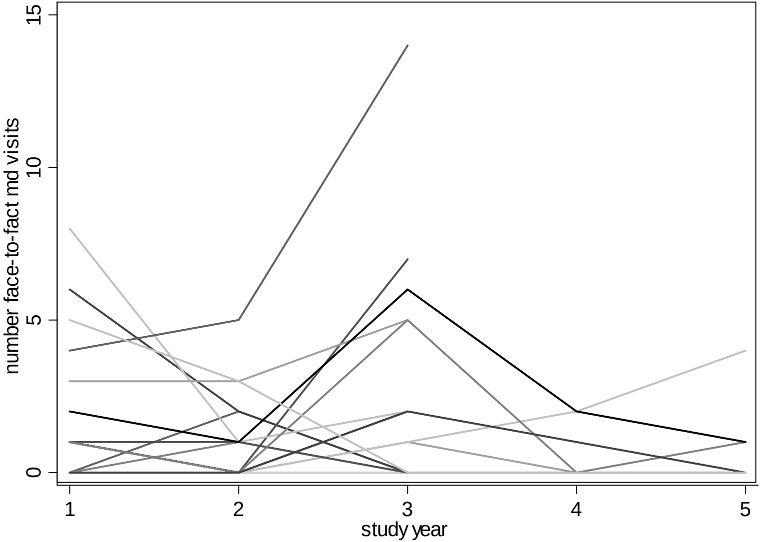
```
. * Year-to-year transitions in doctor visits
. generate mdushort = mdu

. replace mdushort = 4 if mdu >= 4
(4,039 real changes made)

. xttrans mdushort
```

| mdushort | mdushort 0 | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|---|
| 0 | 58.87 | 19.61 | 9.21 | 4.88 | 7.42 | 100.00 |
| 1 | 33.16 | 24.95 | 17.58 | 10.14 | 14.16 | 100.00 |
| 2 | 23.55 | 24.26 | 17.90 | 12.10 | 22.19 | 100.00 |
| 3 | 17.80 | 20.74 | 18.55 | 12.14 | 30.77 | 100.00 |
| 4 | 8.79 | 11.72 | 12.32 | 11.93 | 55.23 | 100.00 |
| Total | 31.81 | 19.27 | 13.73 | 9.46 | 25.73 | 100.00 |

- Time series plots for the first 20 individuals. Much serial correlation.
  - `quietly xtline mdu if _n<=85, overlay legend(off)`



- Because `xtset` set a time variable can use time series commands
  - `Lj.x` gives `x` lagged *j* periods.
- Can compute autocorrelations for a variable.

```
. corr mdu L.mdu L2.mdu L3.mdu
(obs=3,230)

             |                L.       L2.       L3.
             |      mdu      mdu      mdu      mdu
-------------+------------------------------------
         mdu |
         --. |   1.0000
         L1. |   0.6767   1.0000
         L2. |   0.4974   0.6289   1.0000
         L3. |   0.3714   0.4603   0.5994   1.0000
```

- High serial correlation: $\text{Cor}[y_t, y_{t-3}] = 0.37$ and approximately AR(1).
- Note that estimated autocorrelations without imposing stationarity.
- Weakness: Only 3230 observations used as needed `L3.mdu`

- Command pwcorr uses all available data to compute each correlation

```
. pwcorr mdu L.mdu L2.mdu L3.mdu L4.mdu
```

| | mdu | L.mdu | L2.mdu | L3.mdu | L4.mdu |
|---|---|---|---|---|---|
| mdu | 1.0000 | | | | |
| L.mdu | 0.6184 | 1.0000 | | | |
| L2.mdu | 0.4744 | 0.6029 | 1.0000 | | |
| L3.mdu | 0.3714 | 0.4602 | 0.5995 | 1.0000 | |
| L4.mdu | 0.3820 | 0.3702 | 0.5054 | 0.6100 | 1.0000 |

# 3. Panel data models overview: linear model review

- Focus is on model with individual-specific effect

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \quad i = 1, ..., N, t = 1, ..., T.$$

  - So different people have different unobserved intercept $\alpha_i$.

- Goal is to consistently estimate slope parameters $\boldsymbol{\beta}$.
- Focus on short panel with $N \rightarrow \infty$ and $T$ small.
- If $\alpha_i$ is a random effect then
  - pooled OLS with cluster-robust s.e.'s is okay
  - random effects GLS/MLE may be more efficient
- If $\alpha_i$ is a fixed effect, so $\text{Cor}[\alpha_i, \mathbf{x}_{it}] \neq 0$
  - pooled OLS and random effects GLS/MLE are inconsistent
  - need fixed effects: regress $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$
  - or first differences: regress $\Delta y_{it}$ on $\Delta \mathbf{x}_{it}$.

# Panel data overview: poisson model review

- These results carry over qualitatively to Poisson panel regression.
- The Poisson individual effects model specifies conditional mean

$$\begin{aligned} \mathrm{E}[y_{it}|\mathbf{x}_{it},\boldsymbol{\beta},\alpha_i] &= \exp(\delta_i+\mathbf{x}_{it}'\boldsymbol{\beta}) \\ &= \alpha_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta}) \end{aligned}$$

- The effect is multiplicative rather than additive.
- If $\alpha_i$ is a random effect then
  - pooled Poisson with cluster-robust s.e.'s is okay
  - random effects Poisson GLS/MLE may be more efficient
- If $\alpha_i$ is a fixed effect, so $\mathrm{Cov}[\alpha_i,\mathbf{x}_{it}]\neq 0$
  - pooled Poisson and random effects Poisson are inconsistent
  - need fixed effects explained below
  - or quasi first differences

# Static Panel Count Estimators

- Consider five panel Poisson estimators
  - Pooled Poisson with cluster-robust errors
  - Population-averaged Poisson (GEE)
  - Poisson random effects (gamma and normal)
  - Poisson mixed models
  - Poisson fixed effects
- Can additionally apply most of these to negative binomial.

# Panel Poisson method 1: pooled Poisson

- Specify

$$y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta} \sim Poisson[\exp(\mathbf{x}_{it}'\boldsymbol{\beta})]$$

- Pooled Poisson of $y_{it}$ on intercept and $\mathbf{x}_{it}$ gives consistent $\boldsymbol{\beta}$.
  - But get cluster-robust standard errors where cluster on the individual.
  - These control for both overdispersion and correlation over $t$ for given $i$.

- Pooled Poisson with cluster-robust standard errors

```
. * Pooled Poisson estimator with cluster-robust standard errors
. poisson mdu lcoins ndisease female age lfam child, vce(cluster id)

Iteration 0:   log pseudolikelihood = -62580.248
Iteration 1:   log pseudolikelihood = -62579.401
Iteration 2:   log pseudolikelihood = -62579.401

Poisson regression                                Number of obs   =      20186
                                                  Wald chi2(6)    =     476.93
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -62579.401                 Pseudo R2       =     0.0609

                              (Std. Err. adjusted for 5908 clusters in id)
```

| mdu | Coef. | Robust Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| lcoins | -.0808023 | .0080013 | -10.10 | 0.000 | -.0964846 | -.0651199 |
| ndisease | .0339334 | .0026024 | 13.04 | 0.000 | .0288328 | .039034 |
| female | .1717862 | .0342551 | 5.01 | 0.000 | .1046473 | .2389251 |
| age | .0040585 | .0016891 | 2.40 | 0.016 | .000748 | .0073691 |
| lfam | -.1481981 | .0323434 | -4.58 | 0.000 | -.21159 | -.0848062 |
| child | .1030453 | .0506901 | 2.03 | 0.042 | .0036944 | .2023961 |
| _cons | .748789 | .0785738 | 9.53 | 0.000 | .5947872 | .9027907 |

By comparison, the default (non cluster-robust) s.e.'s are 1/4 as large.
$\Rightarrow$ The default (non cluster-robust) t-statistics are 4 times as large!!

# Panel Poisson method 2: population-averaged

- This method is the standard method in statistics.
  - Again assume there is no fixed effects problem.
  - But want more efficient estimator than pooled Poisson.
- Assume that for the $i^{th}$ observation moments are like for GLM Poisson

$$\begin{aligned} \mathrm{E}[y_{it}|\mathbf{x}_{it}] &= \exp(\mathbf{x}_{it}'\boldsymbol{\beta}) \\ \mathrm{V}[y_{it}|\mathbf{x}_{it}] &= \phi \times \exp(\mathbf{x}_{it}'\boldsymbol{\beta}). \end{aligned}$$

- Assume constant correlation between $y_{it}$ and $y_{is}$

$$\mathrm{Cov}[y_{it}, y_{is}|\mathbf{x}_{it}, \mathbf{x}_{is}] = \rho.$$

- Estimate by the generalized estimating equations (GEE) estimator or population-averaged estimator (PA) of Liang and Zeger (1986).
  - This is essentially nonlinear feasible GLS.
- Get a cluster-robust estimate of the variance matrix that is correct even if $\mathrm{Cov}[y_{it}, y_{is}|\mathbf{x}_{it}, \mathbf{x}_{is}] \neq \rho$.

- Population-averaged Poisson with unstructured correlation
  - `xtpoisson mdu lcoins ndisease female age lfam child, ///`
    `pa corr(unstr) vce(robust)`

```
GEE population-averaged model                   Number of obs      =     20186
Group and time vars:              id year       Number of groups   =      5908
Link:                                 log       Obs per group: min =         1
Family:                           Poisson                      avg =       3.4
Correlation:                 unstructured                      max =         5
                                                Wald chi2(6)       =    508.61
Scale parameter:                        1       Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
```

| mdu | Coef. | Semi-robust Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| lcoins | -.0804454 | .0077782 | -10.34 | 0.000 | -.0956904 | -.0652004 |
| ndisease | .0346067 | .0024238 | 14.28 | 0.000 | .0298561 | .0393573 |
| female | .1585075 | .0334407 | 4.74 | 0.000 | .0929649 | .2240502 |
| age | .0030901 | .0015356 | 2.01 | 0.044 | .0000803 | .0060999 |
| lfam | -.1406549 | .0293672 | -4.79 | 0.000 | -.1982135 | -.0830962 |
| child | .1013677 | .04301 | 2.36 | 0.018 | .0170696 | .1856658 |
| _cons | .7764626 | .0717221 | 10.83 | 0.000 | .6358897 | .9170354 |

Generally s.e.'s are within 10% of pooled Poisson cluster-robust s.e.'s.
The default (non cluster-robust) t-statistics are 3.5 − 4 times larger,

- The correlations $\mathrm{Cor}[y_{it}, y_{is}|\mathbf{x}_i]$ for PA (unstructured) are not equal.
  - But they are not declining as fast as AR(1).

```
. matrix list e(R)

symmetric e(R)[5,5]
            c1          c2          c3          c4          c5
r1           1
r2   .53143297           1
r3   .40817495   .58547795           1
r4   .32357326   .35321716   .54321752           1
r5   .34152288   .29803555   .43767583   .61948751           1
```

- Can allow AR(1) correlation so $\mathrm{Cor}(y_{it}, y_{is}|\mathbf{x}_i) = \rho^{|t-s|}$
  - yields similar estimates.

```
. * Allow for AR(1) correlation in y
. xtpoisson mdu lcoins ndisease female age lfam child, pa corr(ar1) vce(robust)

note:  observations not equally spaced
       modal spacing is delta year = 1 unit
       12 groups omitted from estimation
note:  some groups have fewer than 2 observations
       not possible to estimate correlations for those groups
       265 groups omitted from estimation


Iteration 1: tolerance = .01684555
Iteration 2: tolerance = .00011928
Iteration 3: tolerance = 3.741e-07

GEE population-averaged model                   Number of obs      =     19,881
Group and time vars:                 id year    Number of groups   =      5,631
Link:                                    log    Obs per group:
Family:                              Poisson                  min =          2
Correlation:                           AR(1)                  avg =        3.5
                                                              max =          5
                                                Wald chi2(6)       =     478.00
Scale parameter:                           1    Prob > chi2        =     0.0000

                                    (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.0801371   .0079898   -10.03   0.000    -.0957968   -.0644773
    ndisease |   .0337386   .0025236    13.37   0.000     .0287924    .0386848
      female |   .1654749   .0344277     4.81   0.000     .0979978     .232952
         age |   .0040851   .0015731     2.60   0.009     .0010019    .0071683
        lfam |  -.1220327   .0303412    -4.02   0.000    -.1815003   -.0625651
       child |   .0797197   .0443214     1.80   0.072    -.0071486    .1665881
       _cons |   .7358995   .0742563     9.91   0.000     .5903599    .8814391
------------------------------------------------------------------------------
```

# Panel Poisson method 3: random effects

- This method is used more in econometrics.
  - No FE but want more efficient estimator than pooled Poisson.
- Poisson random effects model is

$$y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}, \alpha_i \sim Poiss[\alpha_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta})] \sim Poiss[\exp(\ln \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})]$$

  where $\alpha_i$ is unobserved but is not correlated with $\mathbf{x}_{it}$.
- RE estimator 1: Assume $\alpha_i$ is $Gamma[1, \eta]$ distributed
  - closed-form solution exists (negative binomial)
  - $\mathrm{E}[y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}] = \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$
- RE estimator 2: Assume $\ln \alpha_i$ is $\mathcal{N}[0, \sigma_\varepsilon^2]$ distributed
  - closed-form solution does not exist (one-dimensional integral)
  - can extend to slope coefficients (higher-dimensional integral)
  - $\mathrm{E}[y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}] \neq \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$!!! Now different conditional mean.

- Poisson random effects (gamma) with panel robust se's
  - `xtpoisson mdu lcoins ndisease female age lfam child, re vce(cluster id)`

```
Random-effects Poisson regression               Number of obs     =     20,186
Group variable: id                              Number of groups  =      5,908

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

                                                Wald chi2(6)      =    5407.91
Log pseudolikelihood  = -43240.556              Prob > chi2       =     0.0000

                                 (Std. Err. adjusted for 5,908 clusters in id)
```

| mdu | Coef. | Robust Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| lcoins | -.0878258 | .0079239 | -11.08 | 0.000 | -.1033563 | -.0722952 |
| ndisease | .0387629 | .0024087 | 16.09 | 0.000 | .0340421 | .0434838 |
| female | .1667192 | .0345869 | 4.82 | 0.000 | .09893 | .2345083 |
| age | .0019159 | .0016533 | 1.16 | 0.247 | -.0013244 | .0051563 |
| lfam | -.1351786 | .0360629 | -3.75 | 0.000 | -.2058606 | -.0644966 |
| child | .1082678 | .0533869 | 2.03 | 0.043 | .0036314 | .2129043 |
| _cons | .7574177 | .0831229 | 9.11 | 0.000 | .5944999 | .9203355 |
| /lnalpha | .0251256 | .0905423 | | | -.1523339 | .2025852 |
| alpha | 1.025444 | .092846 | | | .8587015 | 1.224564 |

```
LR test of alpha=0: chibar2(01) = 3.9e+04             Prob >= chibar2 = 0.000
```

The default (non cluster-robust) t-statistics are 2.5 times larger because default do not control for overdispersion.

- Poisson random effects (normal distribution) with panel robust se's
  - computation is fast as one-dimensional integral and ways to speed up
  - `xtpoisson mdu lcoins ndisease female age lfam child, re vce(cluster id)`

```
Random-effects Poisson regression               Number of obs     =     20,186
Group variable: id                              Number of groups  =      5,908

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

Integration method: mvaghermite                 Integration pts.  =         12

                                                Wald chi2(6)      =     798.96
Log likelihood  = -43226.889                    Prob > chi2       =     0.0000
```

| mdu | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| lcoins | -.1145337 | .0072788 | -15.74 | 0.000 | -.1287998 | -.1002676 |
| ndisease | .0408695 | .0022941 | 17.81 | 0.000 | .0363731 | .045366 |
| female | .208436 | .0304848 | 6.84 | 0.000 | .148687 | .268185 |
| age | .002674 | .0011961 | 2.24 | 0.025 | .0003297 | .0050182 |
| lfam | -.1443327 | .0265365 | -5.44 | 0.000 | -.1963432 | -.0923222 |
| child | .0737146 | .0344697 | 2.14 | 0.032 | .0061553 | .141274 |
| _cons | .2872796 | .0641625 | 4.48 | 0.000 | .1615234 | .4130359 |
| /lnsig2u | .0549982 | .0254991 | | | .005021 | .1049755 |
| sigma_u | 1.027881 | .013105 | | | 1.002514 | 1.05389 |

```
LR test of sigma_u=0: chibar2(01) = 3.9e+04           Prob >= chibar2 = 0.000
```

# Panel Poisson method 4: Mixed Models

- Now part or all of $\beta$ is $\beta_i$ normally distributed.
- Stata mixed models here can take a very, very long time !
- Estimation is by quadrature
  - mvaghermite: mean-variance adaptive Gauss-Hermite quadrature
  - mcaghermite: mode-curvature adaptive Gauss-Hermite quadrature
  - ghermite: nonadaptive Gauss-Hermite quadrature
  - laplace: Laplacian approximation (faster)
    - equivalent to mode-curvature adaptive Gaussian quadrature with one integration
- Define
  - `global xlist lcoins ndisease female age lfam child`
- A mixed model with intercept only varying can be estimated several ways
  - `meqrpoisson mdu $xlist || id:`
  - `mepoisson mdu $xlist || id:`
  - `xtpoisson mdu $xlist, re normal vce(cluster id)`
  - `meqrpoisson` takes a long time.

# Mixed models (continued)

- Can also allow a slope parameter to vary
  - `mepoisson mdu $xlist || id: ndisease`
- And also allow slope parameter to covary with intercept parameter
  - `mepoisson mdu $xlist || id: ndisease, cov(unstructured) intpoints(9)`

# Panel Poisson method 5: fixed effects

- Poisson fixed effects model is

$$y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}, \alpha_i \sim Poiss[\alpha_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta})] \sim Poiss[\exp(\ln \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta})]$$

  where $\alpha_i$ is unobserved and is possibly correlated with $\mathbf{x}_{it}$.
- In theory need to estimate $\boldsymbol{\beta}$ and $\alpha_1, ..., \alpha_N$.
  - potential incidental parameters problem $N+K$ parameters and $NT$ observations with $N \rightarrow \infty$.
  - but no problem as can eliminate $\alpha_i$.
- Eliminate $\alpha_i$ by quasi-differencing as follows

$$\begin{array}{rrcll} & \mathrm{E}[y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\alpha_i] &=& \alpha_i\lambda_{it} & \quad \lambda_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta}) \\ \Rightarrow & \mathrm{E}[\bar{y}_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\alpha_i] &=& \alpha_i\bar{\lambda}_i & \quad \bar{\lambda}_i = T^{-1}\sum_t \lambda_{it} \\ \Rightarrow & \mathrm{E}\left[\left(y_{it} - (\lambda_{it}/\bar{\lambda}_i)\bar{y}_i\right)|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}\right] &=& 0 & \end{array}$$

  - The first line assumes regressors $\mathbf{x}_{it}$ are strictly exogenous.
  - This is stronger than weakly exogenous.

- The final result implies

$$\mathrm{E}\left[\mathbf{x}_{it}\left(y_{it}-\frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right)\right]=\mathbf{0}.$$

- Poisson fixed effects estimator solves the corresponding sample moment conditions

$$\textstyle\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\left(y_{it}-\frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right)=\mathbf{0}, \quad \text{where } \lambda_{it}=\exp(\mathbf{x}_{it}'\boldsymbol{\beta}).$$

  - Get cluster-robust standard errors
  - Bootstrap `xtpoisson, re` or use add-on `xtpqml`.

- Consistency requires

$$\mathrm{E}[y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\alpha_i]=\alpha_i\exp(\mathbf{x}_{it}'\boldsymbol{\beta}).$$

- Poisson fixed effects with panel bootstrap se's
  - `xtpoisson mdu lcoins ndisease female age lfam child, fe vce(robust)`

```
Conditional fixed-effects Poisson regression    Number of obs      =     17,791
Group variable: id                              Number of groups   =      4,977

                                                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)
```

| mdu | Coef. | Robust 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 |

The default (non cluster-robust) t-statistics are 2 times larger.

- Remarkably the Poisson FE estimator for $\boldsymbol{\beta}$ can also be obtained in the following ways under fully parametric assumption that

$$y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}, \alpha_i \sim Poiss[\alpha_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta})]$$

- 1. Obtain the MLE of $\boldsymbol{\beta}$ and $\alpha_1, ..., \alpha_N$.
- 2. Obtain the conditional MLE based on the conditional density

$$f(y_{i1}, ..., y_{it}|\bar{y}_i, \mathbf{x}_{i1}, ..., \mathbf{x}_{iT}, \boldsymbol{\beta}, \alpha_i) = \frac{\prod_{t=1}^{T} f(y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}, \alpha_i)}{f(\bar{y}_i|\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}, \boldsymbol{\beta}, \alpha_i)}$$

- But should then use cluster-robust standard errors and not default ML se's.

- Strength of fixed effects versus random effects
  - Allows $\alpha_i$ to be correlated with $\mathbf{x}_{it}$.
  - So consistent estimates if regressors are correlated with the error provided regressors are correlated only with the time-invariant component of the error
  - An alternative to IV to get causal estimates.
- Limitations:
  - Coefficients of time-invariant regressors are not identified
  - For identified regressors standard errors can be much larger
  - Marginal effect in a nonlinear model depend on $\alpha_i$

$$\mathrm{ME}_j = \partial \mathrm{E}[y_{it}]/\partial \mathbf{x}_{it,j} = \alpha_i \exp(\mathbf{x}_{it}'\boldsymbol{\beta})\beta_j$$

and $\alpha_i$ is unknown.

# Panel Poisson: estimator comparison

- Compare following estimators
  - pooled Poisson with cluster-robust s.e.'s
  - pooled population averaged Poisson with unstructured correlations and cluster-robust s.e.'s
  - random effects Poisson with gamma random effect and cluster-robust s.e.'s
  - random effects Poisson with normal random effect and default s.e.'s
  - fixed effects Poisson and cluster-robust s.e.'s
- Find that
  - similar results for all but FE
  - note that these data are not good to illustrate FE as regressors have little within variation.

```
* Comparison of Poisson panel estimators
* using panel-robust standard errors
global xlist lcoins ndisease female age lfam child
quietly poisson mdu $xlist, vce(cluster id)
estimates store POOLED
quietly xtpoisson mdu $xlist, pa corr(unstr) vce(robust)
estimates store POPAVE
quietly xtpoisson mdu $xlist, re vce(cluster id)
estimates store RE_GAMMA
quietly xtpoisson mdu $xlist, re normal vce(cluster id)
estimates store RE_NORMAL
quietly xtpoisson mdu $xlist, fe vce(robust)
estimates store FIXED
estimates table POOLED POPAVE RE_GAMMA RE_NORMAL FIXED, ///
 equations(1) b(%8.4f) se stats(N ll) stfmt(%8.0f)
```

- Compare different Poisson panel estimators with panel-robust s.e.'s

| Variable | POOLED | POPAVE | RE_GAMMA | RE_NOR~L | FIXED |
|---|---|---|---|---|---|
| #1 | | | | | |
| lcoins | -0.0808 | -0.0804 | -0.0878 | -0.1145 | |
| | 0.0080 | 0.0078 | 0.0079 | 0.0072 | |
| ndisease | 0.0339 | 0.0346 | 0.0388 | 0.0409 | |
| | 0.0026 | 0.0024 | 0.0024 | 0.0023 | |
| female | 0.1718 | 0.1585 | 0.1667 | 0.2084 | |
| | 0.0343 | 0.0334 | 0.0346 | 0.0310 | |
| age | 0.0041 | 0.0031 | 0.0019 | 0.0027 | -0.0112 |
| | 0.0017 | 0.0015 | 0.0017 | 0.0017 | 0.0091 |
| lfam | -0.1482 | -0.1407 | -0.1352 | -0.1443 | 0.0877 |
| | 0.0323 | 0.0294 | 0.0361 | 0.0359 | 0.1161 |
| child | 0.1030 | 0.1014 | 0.1083 | 0.0737 | 0.1060 |
| | 0.0507 | 0.0430 | 0.0534 | 0.0534 | 0.0786 |
| _cons | 0.7488 | 0.7765 | 0.7574 | 0.2873 | |
| | 0.0786 | 0.0717 | 0.0831 | 0.0829 | |
| lnalpha | | | | | |
| _cons | | | 0.0251 | | |
| | | | 0.0905 | | |
| lnsig2u | | | | | |
| _cons | | | | 0.0550 | |
| | | | | 0.0271 | |
| Statistics | | | | | |
| N | 20186 | 20186 | 20186 | 20186 | 17791 |
| ll | -62579 | | -43241 | -43227 | -24173 |

legend: b/se

# Panel negative binomial

- Fixed and random effects for negative binomial also exist.
    - But efficiency gains may not be great
    - and fixed effects negative binomial is not really fixed effects.
- Simplest to work with Poisson
    - but make sure get cluster-robust standard errors.

# 5. Dynamic panel count estimators

- Now include lagged $y's$ as regressors
- Let $\mu_{\iota t} = \mathrm{E}[y_{it}|\alpha_i, y_{i,t-1}, y_{i,t-2}, \mathbf{x}_{it}, \mathbf{x}_{it-1}, ...]$
- It is not clear how to include the lagged $y_{it}'s$.
- We use the exponential feedback model

$$\mu_{\iota t} = \alpha_i \exp(\rho y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta})$$

  but this is potentially explosive.
- A better model may be the linear feedback model:

$$\mu_{\iota t} = \alpha_i \{\rho y_{i,t-1} + \exp(\mathbf{x}_{it}'\boldsymbol{\beta})\}$$

# Non-Fixed effects Models

- We can estimate this model in various ways already seen.
  - pooled
  - random effects.
- The correlated random effects model is a random effects model that adds $y_{i0}$ and $\bar{\mathbf{x}}_i$ as controls
  - $\mu_{\iota t} = \alpha_i \exp(\rho y_{i,t-1} + \mathbf{x}'_{it}\boldsymbol{\beta})$ where $\alpha_i = \exp(\delta_0 y_{i0} + \bar{\mathbf{x}}'_i \boldsymbol{\lambda} + \varepsilon_i)$
  - this is getting close to fixed effects.

# Linear Dynamic Panel Models with Individual Effects

- Before considering dynamic count models with individual fixed effects consider the linear model.
- **Leading example**: AR(1) model with individual specific effects

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}.$$

- Four reasons for $y_{it}$ being serially correlated over time:
  - **Unobserved heterogeneity:** via $\alpha_i$ (permanent)
  - **True state dependence:** via $y_{i,t-1}$ (dampens over time)
  - **Observed heterogeneity:** via $\mathbf{x}_{it}$ which may be serially correlated
  - **Error correlation:** via $\varepsilon_{it}$
- There is a literature on dynamic models with random effects.
- We focus on case where $\alpha_i$ is a **fixed effect**
  - and $T$ small so cannot consistently estimate $\alpha_i$
  - no problem if $T \rightarrow \infty$ and $\varepsilon_{it}$ serially uncorrelated.

# Fixed Effects Estimator is Inconsistent

- Mean difference yields

$$(y_{it} - \bar{y}_i) = \gamma(y_{i,t-1} - \bar{y}_{i,-1}) + (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i).$$

- Problem: regressor $(y_{i,t-1} - \bar{y}_{i,-1})$ correlated with $(\varepsilon_{it} - \bar{\varepsilon}_i)$
  - since $y_{it}$ (part of $\bar{y}_{i,-1}$) is correlated with $\varepsilon_{it}$
  - and $y_{i,t-1}$ is correlated with $\varepsilon_{it}$ (part of $\bar{\varepsilon}_i$).
- This inconsistency is called Nickell bias
  - inconsistency is $O(T^{-1})$.

# Arellano-Bond Estimator

- **First-difference** to eliminate $\alpha_i$ (rather than mean-difference)

$$(y_{it} - y_{i,t-1}) = \gamma(y_{i,t-1} - y_{i,t-2}) + (\mathbf{x}_{it} - \mathbf{x}'_{i,t-1})\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}).$$

- **OLS inconsistent** as $(y_{i,t-1} - y_{i,t-2})$ correlated with $(\varepsilon_{it} - \varepsilon_{i,t-1})$ (even under assumption $\varepsilon_{it}$ is serially uncorrelated).
- But $y_{i,t-2}$ is uncorrelated with $(\varepsilon_{it} - \varepsilon_{i,t-1})$, **if $\varepsilon_{it}$ serially uncorrelated**
  so can use $y_{i,t-2}$ as an **instrument** for $(y_{i,t-1} - y_{i,t-2})$.
- Arellano-Bond is a variation that uses **unbalanced set** of instruments with further lags as instruments.
  For $t = 3$ can use $y_{i1}$, for $t = 4$ can use $y_{i1}$ and $y_{i2}$, and so on.
- Stata commands
  - `xtabond` for Arellano-Bond
  - `xtdpdsys` for Blundell-Bond (more efficient than `xtabond`)
  - `xtdpd` for more complicated models than `xtabond` and `xtdpdsys`.

# Example Stata for linear Arellano Bond

```
.   * Optimal or two-step GMM for a dynamic panel model
.   xtabond lwage occ south smsa ind, lags(2) maxldep(3) ///
     pre(wks,lag(1,2)) endogenous(ms,lag(0,2)) ///
.   endogenous(union,lag(0,2)) twostep vce(robust) ///
.   artests(3)
.   * Test whether error is serially correlated
.   estat abond
.   * Test of overidentifying restrictions
.   estat sargan
.   * Arellano/Bover or Blundell/Bond for dynamic panel
.   xtdpdsys lwage occ south smsa ind, lags(2) ///
     maxldep(3) pre(wks,lag(1,2)) endog(ms,lag(0,2)) ///
     endogenous(union,lag(0,2)) twostep vce(robust) ///
.   artests(3)
```

# Count dynamic panel estimators: Fixed effects

- Extend Arellano-Bond setup for linear model to nonlinear model.
- Sequential moment conditions

$$\mathrm{E}[y_{it}|y_{it-1},...,y_{i1},\mathbf{x}_{it},\ldots,\mathbf{x}_{i1},\alpha_i]=\alpha_i\lambda_{it}$$

  - where we use $\alpha_i\lambda_{it}=\alpha_i\exp(\rho y_{i,t-1}+\mathbf{x}_{it}'\boldsymbol{\beta})$.
- Then $\mathrm{E}[\frac{\lambda_{it-1}}{\lambda_{it}}y_{it}|y_{it-1},...,y_{i1},\mathbf{x}_{it},\ldots,\mathbf{x}_{i1},\alpha_i]=\alpha_i\lambda_{it-1}$.
- Also $\mathrm{E}[y_{it-1}|y_{it-2},...,y_{i1},\mathbf{x}_{it-1},\ldots,\mathbf{x}_{i1},\alpha_i]=\alpha_i\lambda_{it-1}$.
- So subtracting

$$\mathrm{E}[(\lambda_{i,t-1}/\lambda_{it})y_{it-1}-y_{it}|y_{it-1},...,y_{i1},\mathbf{x}_{it},\ldots,\mathbf{x}_{i1}]=0.$$

- So can do GMM based on moment conditions

$$\mathrm{E}[\mathbf{z}_{it}\{y_{it}-(\lambda_{i,t-1}/\lambda_{it})y_{it-1}\}]=\mathbf{0}$$

  - where $\mathbf{z}_{it}=(y_{it-1},\mathbf{x}_{it})$ in just-identified case.
- These can be coded up in Stata using the `gmm` command.

# Example count dynamic model with fixed effects

- Patents, lagged patents and current and lagged R&D spending.

```
. * Variable descriptions and summary statistics
. use racd09data.dta, clear

. global XLISTD PAT1 LOGR LOGR1 LOGR2 LOGK SCISECT dyear2 dyear3 dyear4 dyear5

. describe PAT $XLISTD

              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------------------------
PAT             float   %9.0g                 Number of (successful) patents applied for this
                                                year
PAT1            float   %9.0g                 Number of (successful) patents applied for lagged
                                                one year
LOGR            float   %9.0g                 Logarithm of R&D spending this year (in 1972$)
LOGR1           float   %9.0g                 Logarithm of R&D spending lagged one year (in
                                                1972$)
LOGR2           float   %9.0g                 Logarithm of R&D spending lagged two years (in
                                                1972$)
LOGK            float   %9.0g                 Logarithm of the book value of capital in 1972
SCISECT         float   %9.0g                 Equals 1 if firm in the scientific sector
dyear2          float   %9.0g                 = 1 if YEAR = 2
dyear3          float   %9.0g                 = 1 if YEAR = 3
dyear4          float   %9.0g                 = 1 if YEAR = 4
dyear5          float   %9.0g                 = 1 if YEAR = 5

. summarize PAT1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        PAT1 |      1,730    35.87341    72.76243          0        528
```

# Program for count dynamic FE

- This program to be used by Stata gmm command.

```
. program drop gmm_poipre

. do "C:\Users\ccameron\AppData\Local\Temp\STD2f80_000000.tmp"

. * Fixed effects GMM using Chamberlain transformation
. * This program is the same as gmm_poipre in Stata manual [r]gmm
. program gmm_poipre
  1.     version 11
  2.     syntax varlist if, at(name) myrhs(varlist) ///
>     mylhs(varlist) myidvar(varlist)
  3.     quietly {
  4.     tempvar mu mubar ybar
  5.     gen double `mu' = 0 `if'
  6.     local j = 1
  7.     foreach var of varlist `myrhs' {
  8.        replace `mu' = `mu' + `var'*`at'[1,`j'] `if'
  9.        local j = `j' + 1
 10.        }
 11.     replace `mu' = exp(`mu')
 12.     replace `varlist' = L.`mylhs' - L.`mu'*`mylhs'/`mu' `if'
 13.     }
 14. end
```

# Setup and command

- Just identified example.

```
Step 1
Iteration 0:   GMM criterion Q(b) =  13.821058
Iteration 1:   GMM criterion Q(b) =  .96221647
Iteration 2:   GMM criterion Q(b) =  .00240722
Iteration 3:   GMM criterion Q(b) =  2.011e-08
Iteration 4:   GMM criterion Q(b) =  9.320e-20

note: model is exactly identified

GMM estimation

Number of parameters =   7
Number of moments    =   7
Initial weight matrix: Unadjusted                Number of obs   =      1,384

                              (Std. Err. adjusted for 346 clusters in id)
```

| | Coef. | Robust Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| /PAT1 | .0021701 | .0011062 | 1.96 | 0.050 | 2.12e-06 | .0043381 |
| /LOGR | -.62016 | 1.356658 | -0.46 | 0.648 | -3.279161 | 2.038841 |
| /LOGR1 | -.100405 | .127623 | -0.79 | 0.431 | -.3505415 | .1497316 |
| /LOGR2 | .0315067 | .0896787 | 0.35 | 0.725 | -.1442603 | .2072738 |
| /dyear3 | .0576212 | .0862853 | 0.67 | 0.504 | -.1114948 | .2267372 |
| /dyear4 | .0028215 | .1792061 | 0.02 | 0.987 | -.348416 | .354059 |
| /dyear5 | .0723403 | .3120826 | 0.23 | 0.817 | -.5393303 | .6840108 |

```
Instruments for equation 1: PAT2 LOGR1 LOGR2 LOGR3 dyear3 dyear4 dyear5
```

# References for count panel


- Chapter 9 of A. Colin Cameron and Pravin K. Trivedi (2013)
  *Regression Analysis of Count Data (RACD)*
  Second edition, Cambridge Univ. Press.
- A. Colin Cameron and Pravin K. Trivedi (2013)
  "Count Panel Data," in B. Baltagi ed., Oxford Handbook of Panel Data, 2015, Oxford: Oxford University Press, pp.233-256.