Day 1B
Count Data Regression: Part 2

A. Colin Cameron
Univ. of Calif. - Davis
... for
Center of Labor Economics
Norwegian School of Economics
Advanced Microeconometrics

Aug 28 - Sept 1, 2017

1. Introduction

- Count data models are for dependent variable $y = 0, 1, 2, ...$
- Previously considered basic cross-section
  - Poisson, negative binomial, GLM's
  - Richer models: hurdle, zero-inflated, ..
- Now consider
  - mixture models
  - endogenous regressors
  - panel data.
- Many of these methods generalize to other nonlinear models.

Outline

2. Finite mixtures model

- Density is weighted sum of two (or more) densities
  - Permits flexible models e.g. bimodal from Poissons.

- For an m-component model

$$f(y|\mathbf{x},\boldsymbol{\theta},\boldsymbol{\pi}) = \sum_{j=1}^{m} \pi_j f_j(y|\mathbf{x},\boldsymbol{\theta}_j), \qquad 0 \leq \pi_j \leq 1, \ \sum_{j=1}^{m} \pi_j = 1.$$

- For a 2-component model

$$f(y|\mathbf{x},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\pi) = \pi f_1(y|\mathbf{x},\boldsymbol{\theta}_1) + (1-\pi) f_2(y|\mathbf{x},\boldsymbol{\theta}_2)$$

- MLE maximizes

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln(\pi f_1(y|\mathbf{x},\boldsymbol{\theta}_1) + (1-\pi) f_2(y|\mathbf{x},\boldsymbol{\theta}_2)).$$

  - Can restrict some parameters to be the same. e.g. only intercept differs
  - EM algorithm often used rather than Newton-Raphson.

- Determining the number of components is a nonstandard inference problem as testing at boundary of parameter space.
  - Simple approach is to use BIC or CAIC.
  - Or do appropriate bootstrap for the likelihood ratio test.
- An alternative to MLE is minimum Hellinger distance estimation.

$$d(\boldsymbol{\theta}) = \sum_{k=0}^{\infty} \left[(\overline{p}_k)^{1/2} - \left(\frac{1}{N} \sum_{i=1}^{N} f(y_i = k | \mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\pi}) \right)^{1/2} \right]^2$$

  - where $\overline{p}_k$ equals fraction of observations with $y_i = k$.
  - attraction is that it is less influenced by outlying observations
  - estimate using an iterative method (HELMIX)

Latent class model

- Finite mixture model can be interpreted as a latent class model.
- There are two types of people (given observables **x**)
  - e.g. "sick" type and "healthy" type
  - there is a probability of being drawn from either type.
- Similar to unobserved heterogeneity in duration data models.
- Stata version 15 command
  - `fmm 2, vce(robust): poisson docvis $xlist`
- Before version 15 there was user-written command fmm

```
Finite mixture model                              Number of obs     =      3,677
Log pseudolikelihood = -12100.185
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.Class	(base outcome)					
2.Class						
_cons	-.5980831	.1171272	-5.11	0.000	-.8276481	-.368518

```
Class          : 1
Response       : docvis
Model          : poisson
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
docvis						
private	.2393558	.0695013	3.44	0.001	.1031357	.3755759
medicaid	.0463821	.0884509	0.52	0.600	-.1269785	.2197427
age	-.6233526	.1367728	-4.56	0.000	-.8914223	-.3552829
age2	.0045366	.0009492	4.78	0.000	.0026762	.0063971
educyr	.0284599	.0078417	3.63	0.000	.0130905	.0438294
actlim	.1723268	.0733314	2.35	0.019	.0285999	.3160537
totchr	.3286694	.0215553	15.25	0.000	.2864218	.3709169
_cons	21.35464	4.881451	4.37	0.000	11.78717	30.92211

```
Class          : 2
Response       : docvis
Model          : poisson
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
docvis						
private	.1566873	.0656687	2.39	0.017	.0279789	.2853957
medicaid	.1924436	.1383488	1.39	0.164	-.078715	.4636022
age	1.232368	.112787	10.93	0.000	1.011309	1.453426
age2	-.0085471	.0007425	-11.51	0.000	-.0100024	-.0070917
educyr	.0219929	.0084774	2.59	0.009	.0053775	.0386082
actlim	.1486859	.0825088	1.80	0.072	-.0130284	.3104003
totchr	.1898829	.03185	5.96	0.000	.127458	.2523078
_cons	-42.46506	4.251345	-9.99	0.000	-50.79755	-34.13258

Log-likelihood comparison across models:
Poisson -15019; 2-component Poisson -11052; 2-component NB2 -10534;
2-component NB1 -10493.
Last is almost exactly same as hurdle NB and ZINB (-10493).

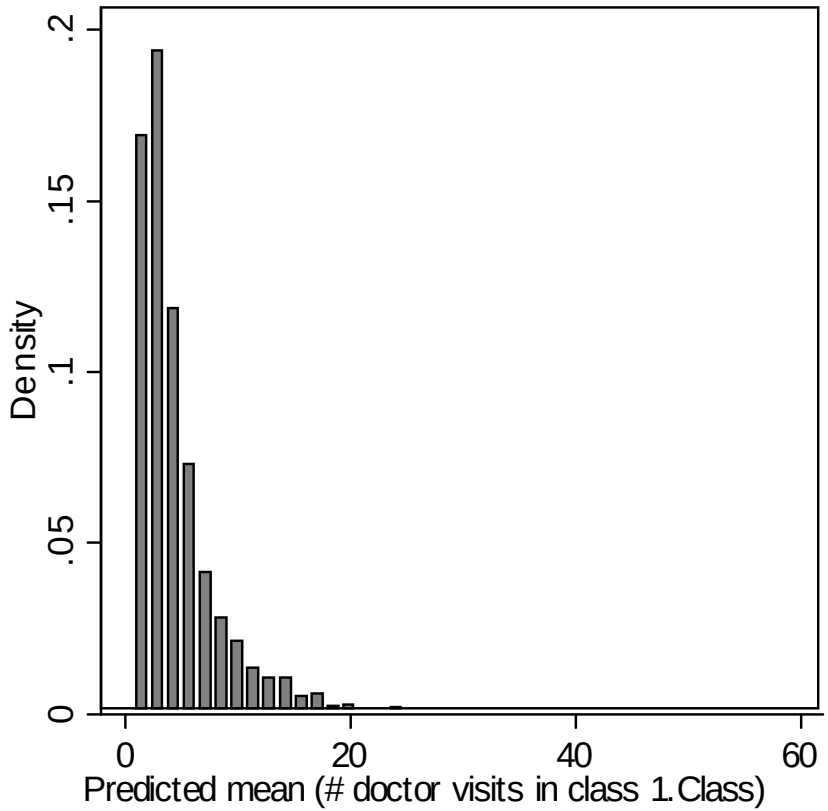
Component 1 occurs with probability 0.65 and is low use.
Component 2 occurs with probability 0.35 and is high use.

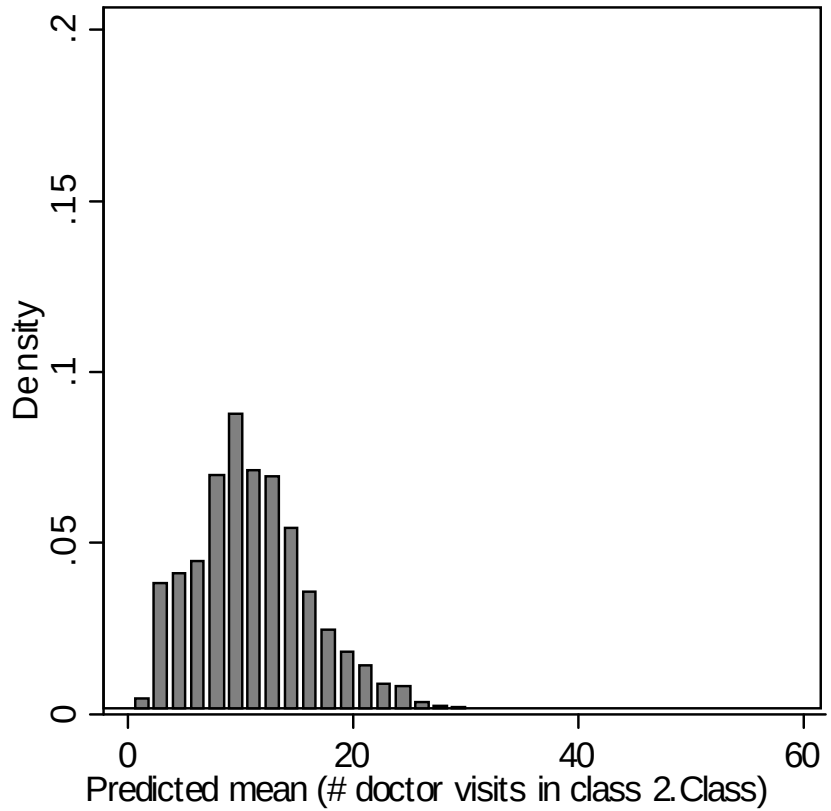
```
. * Obtain latent class probabilities
. estat lcprob

Latent class marginal probabilities             Number of obs     =      3,677
```

	Delta-method Margin	Std. Err.	[95% Conf.	Interval]
Class				
1	.6452176	.0268118	.5911008	.6958574
2	.3547824	.0268118	.3041426	.4088992

```
.
. * Obtain predicted mean for each individual by component
. predict yfit*
(option mu assumed)

. summarize yfit1 yfit2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
yfit1	3,677	5.050474	4.269029	.9507732	50.52482
yfit2	3,677	11.65096	5.670928	.6697477	58.63944


Density
0
.05
.1
.15
.2
0
20
40
60
Predicted mean (# doctor visits in class 1.Class)

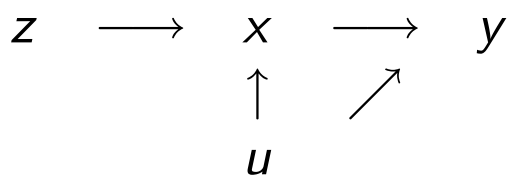

Density
0
.05
.1
.15
.2
0
20
40
60
Predicted mean (# doctor visits in class 2.Class)

3. Endogenous regressor: linear model

- Begin with review of the linear regression model: $y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i$.
- If regressors are correlated with error then OLS is inconsistent.
  - Reason: OLS $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$ so

$$\begin{aligned} \operatorname{plim}\widehat{\boldsymbol{\beta}} &= \boldsymbol{\beta} + (\operatorname{plim} N^{-1}\mathbf{X}'\mathbf{X})^{-1} \operatorname{plim} N^{-1}\mathbf{X}'\mathbf{u} \\ &\neq \boldsymbol{\beta} \text{ if } \operatorname{plim} N^{-1}\mathbf{X}'\mathbf{u} \neq \mathbf{0}. \end{aligned}$$

- Solution: Assume the existence of an instrument z where
  - changes in z are associated with changes in x
  - but changes in z do not led to change in y (aside from indirectly via x)



- Leads to instrumental variables (IV) estimator and two-stage least squares (2SLS) estimator.

- Formally key assumption is:

$$\mathrm{E}[u_i|\mathbf{z}_i] = 0$$

- Just-identified case (# instruments = # endogenous)

$$\widehat{\boldsymbol{\beta}}_{\mathrm{IV}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}.$$

- Over-identified case (# instruments > # endogenous)

$$\widehat{\boldsymbol{\beta}}_{\mathrm{2SLS}} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}.$$

- Example: log-earnings (y) regressed on years of school (x)
  - ability is an omitted regressor so part of error (u) and clearly correlated with x
  - instrument z is correlated with years of school but not directly with earnings
  - example of z may be distance from school or college.

Several Interpretations of Linear IV/2SLS

- 1. Method of Moments
  - $E[u_i|\mathbf{z}_i] = 0 \Rightarrow E[\mathbf{z}_i u_i] = \mathbf{0} \Rightarrow E[\mathbf{z}_i(y_i - \mathbf{x}_i'\boldsymbol{\beta})] = \mathbf{0}$.
  - IV solves corresponding sample moment condition $\sum_{i=1}^{N} \mathbf{z}_i(y_i - \mathbf{x}_i'\boldsymbol{\beta}) = \mathbf{0}$
  - And if overidentified do generalized method of moments (GMM).
- 2. Control Function
  - add predicted residual to control for endogeneity.
  - OLS with additional regressor the residual from first-stage OLS regression of y_2 on all exogenous regressors.
- 3. Two-Stage Least Squares
  - OLS with endogenous regressor y_2 replaced by its predicted value $\widehat{y}_2$ from first-stage OLS regression of y_2 on all exogenous regressors.
- Only methods 1 and 3 extend to nonlinear models such as Poisson.

Poisson endogenous method 1: nonlinear GMM

- Problem is
$$\mathrm{E}[(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))|\mathbf{x}_i] \neq \mathbf{0}.$$
- Assume existence of instruments $\mathbf{z}_i$ such that
$$\begin{aligned} & \mathrm{E}[(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))|\mathbf{z}_i] = \mathbf{0} \\ \Rightarrow \quad & \mathrm{E}[\mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))] = \mathbf{0} \end{aligned}$$

- Just-identified case: $\widehat{\boldsymbol{\beta}}_{\mathrm{MM}}$ solves
$$\sum_{i=1}^{n}(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{z}_i = \mathbf{0}.$$

- Over-identified case $\widehat{\boldsymbol{\beta}}_{\mathrm{GMM}}$ minimizes
$$\left(\sum_{i=1}^{n}(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{z}_i\right)' \mathbf{W} \left(\sum_{i=1}^{n}(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{z}_i\right)$$
  - usually $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$ (called nonlinear 2SLS).

- Literature exists on weighting matrix $\mathbf{W}$ and whether to use different moment condition such as

$$\mathsf{E}\left[\frac{(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{z}_i\right] = \mathbf{0}$$

  - Mullahy (1997), Windmeijer and Santos Silva (1997), Windmeijer (2008).
- We use the simpler $\sum_{i=1}^{n}(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{z}_i = \mathbf{0}$.
- The following output was obtained using own code written using Mata

NL2SLS: Example with `private` (private insurance) endogenous
Instruments are `income` and `ssiratio` (soc sec income / total income)
Estimate by nonlinear 2SLS:

```
. ereturn display
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
private	.5920658	.3401151	1.74	0.082	-.0745475	1.258679
medicaid	.3186961	.1912099	1.67	0.096	-.0560685	.6934607
age	.3323219	.0706128	4.71	0.000	.1939233	.4707205
age2	-.002176	.0004648	-4.68	0.000	-.003087	-.001265
educyr	.0190875	.0092318	2.07	0.039	.0009935	.0371815
actlim	.2084997	.0434233	4.80	0.000	.1233916	.2936079
totchr	.2418424	.013001	18.60	0.000	.2163608	.267324
cons	-11.86341	2.735737	-4.34	0.000	-17.22535	-6.50146

`private` was 0.142 (0.036) and is now 0.592 (0.340)
standard errors much larger with IV
Also `medicaid` changes a lot. Others change little.

Poisson endogenous method 2: control function

- Add error in Poisson model (allows for overdispersion and endogeneity)

$$\begin{aligned}
&\text{Structural eqn:} && y_{1i} \sim \text{Poisson}[\mu_i = \exp(\beta_1 y_{2i} + \mathbf{z}_{1i}'\boldsymbol{\beta}_2 + u_{1i})] \\
&\text{Reduced-form eqn:} && y_{2i} = \gamma_1 z_{2i} + \mathbf{z}_{1i}'\boldsymbol{\gamma}_2 + v_{2i} \\
&\text{Error model:} && u_{1i} = \alpha v_{2i} + \varepsilon_i
\end{aligned}$$

- Then

$$\begin{aligned}
\mu_i|y_{2i}, \mathbf{z}_{1i}, v_{2i}, \varepsilon_i &= \exp(\beta_1 y_{2i} + \mathbf{z}_{1i}'\boldsymbol{\beta}_2 + \alpha v_{2i} + \varepsilon_i) \\
&= \exp(\varepsilon_i)\exp(\beta_1 y_{2i} + \mathbf{z}_{1i}'\boldsymbol{\beta}_2 + \alpha v_{2i}) \\
\mu_i|y_{2i}, \mathbf{z}_{1i}, v_{2i} &= \mathrm{E}[\exp(\varepsilon_i)]\exp(\beta_1 y_{2i} + \mathbf{z}_{1i}'\boldsymbol{\beta}_2 + \alpha v_{2i}) \\
&= \exp(\beta_1 y_{2i} + \mathbf{z}_{1i}'\boldsymbol{\beta}_2 + \alpha v_{2i})
\end{aligned}$$

  where if ε_i is i.i.d. then $\mathrm{E}[\exp(\varepsilon_i)]$ is a constant that is absorbed in $\boldsymbol{\beta}_2$.

- Control function approach
  - 1. OLS of y_2 on z_2 and $\mathbf{z}_1$ gives residual $\widehat{v}_{2i} = y_{2i} - \widehat{\gamma}_1 z_{1i} - \mathbf{z}_{2i}'\widehat{\boldsymbol{\gamma}}_2$.
  - 2. Poisson of y_{1i} on y_{2i}, $\mathbf{z}_{1i}$ and $\widehat{v}_{2i}$ gives IV estimate.

Control function approach for same example.
First-stage: OLS for reduced form

```
. global xlist2 medicaid age age2 educyr actlim totchr

. regress private $xlist2 income ssiratio, vce(robust)

Linear regression                               Number of obs =    3677
                                                F(  8,  3668) =  249.61
                                                Prob > F      =  0.0000
                                                R-squared     =  0.2108
                                                Root MSE      =  .44472
```

private	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
medicaid	-.3934477	.0173623	-22.66	0.000	-.4274884	-.3594071
age	-.0831201	.0293734	-2.83	0.005	-.1407098	-.0255303
age2	.0005257	.0001959	2.68	0.007	.0001417	.0009098
educyr	.0212523	.0020492	10.37	0.000	.0172345	.02527
actlim	-.0300936	.0176874	-1.70	0.089	-.0647718	.0045845
totchr	.0185063	.005743	3.22	0.001	.0072465	.0297662
income	.0027416	.0004736	5.79	0.000	.0018131	.0036702
ssiratio	-.0647637	.0211178	-3.07	0.002	-.1061675	-.0233599
_cons	3.531058	1.09581	3.22	0.001	1.3826	5.679516

Second stage: Poisson with first-stage predicted residual as regressor

```
. predict lpuhat, residual

.
. * Second-stage Poisson with robust SEs
. poisson docvis private $xlist2 lpuhat, vce(robust) nolog

Poisson regression                              Number of obs   =       3677
                                                Wald chi2(8)    =     718.87
                                                Prob > chi2     =     0.0000
Log pseudolikelihood = -15010.614               Pseudo R2       =     0.1303
```

docvis	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
private	.5505541	.2453175	2.24	0.025	.0697407	1.031368
medicaid	.2628822	.1197162	2.20	0.028	.0282428	.4975217
age	.3350604	.0696064	4.81	0.000	.1986344	.4714865
age2	-.0021923	.0004576	-4.79	0.000	-.0030893	-.0012954
educyr	.018606	.0080461	2.31	0.021	.002836	.034376
actlim	.2053417	.0414248	4.96	0.000	.1241505	.286533
totchr	.24147	.0129175	18.69	0.000	.2161523	.2667878
lpuhat	-.4166838	.249347	-1.67	0.095	-.9053949	.0720272
_cons	-11.90647	2.661445	-4.47	0.000	-17.1228	-6.69013

`private` is 0.551 (0.245) compared to (0.340) for NL2SLS

Should bootstrap to get correct s.e.'s (lpuhat is a generated regressor)

```
. * Program and bootstrap for Poisson two-step estimator
. program endogtwostep, eclass
  1.    version 10.1
  2.    tempname b
  3.    capture drop lpuhat2
  4.    regress private $xlist2 income ssiratio
  5.    predict lpuhat2, residual
  6.    poisson docvis private $xlist2 lpuhat2
  7.    matrix `b' = e(b)
  8.    ereturn post `b'
  9. end

. bootstrap _b, reps(400) seed(10101) nodots nowarn: endogtwostep

Bootstrap results                                 Number of obs     =      3,677
                                                  Replications      =        400
```

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
private	.5505541	.2799587	1.97	0.049	.0018452	1.099263
medicaid	.2628822	.1284541	2.05	0.041	.0111169	.5146476
age	.3350604	.0744832	4.50	0.000	.189076	.4810449
age2	-.0021923	.0004888	-4.48	0.000	-.0031504	-.0012342
educyr	.018606	.0091275	2.04	0.042	.0007164	.0364957
actlim	.2053417	.0435134	4.72	0.000	.1200571	.2906264
totchr	.24147	.0134558	17.95	0.000	.2150971	.267843
lpuhat2	-.4166838	.2827737	-1.47	0.141	-.9709101	.1375424
_cons	-11.90647	2.851588	-4.18	0.000	-17.49548	-6.317456

Here little change in standard errors.

Poisson endogenous method 3: structural approach

- Example with binary endogenous regressor y_{2i} is

$$\begin{aligned} &\text{Outcome eqn:} && y_{1i} \sim \text{Poisson}[\mu_i = \exp(\beta_1 y_{2i} + \mathbf{z}_{1i}'\boldsymbol{\beta}_2 + \delta_1 u_i)] \\ &\text{Participation eqn:} && \Pr[y_{2i} = 1] = \Lambda(\mathbf{z}_{2i}'\boldsymbol{\beta}_2 + \lambda_1 u_i) \\ &\text{Error model:} && u_i \sim \mathcal{N}[0,1] \end{aligned}$$

- Estimate using Stata 15 command `gsem`
  - `global xlist2 medicaid age age2 educyr actlim totchr`
  - `gsem (docvis <- private $xlist2 L, poisson) ///`
  - `(private <- $xlist2 income ssiratio L, logit), var(L@1)`
- Can also extend the two-part (hurdle) model to incorporate selection
  - This allows for correlation due to unobservables between process for $y = 0$ or not and process for positives.
  - Terza (1998).

4. Panel data: 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 α_i.

- Goal is to consistently estimate slope parameters $\boldsymbol{\beta}$.
- Focus on short panel with $N \longrightarrow \infty$ and T small.
- If α_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 α_i is a fixed effect, so $\text{Cov}[\alpha_i, \mathbf{x}_{it}] \neq 0$
  - 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 Δy_{it} on $\Delta \mathbf{x}_{it}$.

Panel data: poisson model review

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

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

- The effect is multiplicative rather than additive.
- If α_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 α_i is a fixed effect, so $\mathsf{Cov}[\alpha_i, \mathbf{x}_{it}] \neq 0$
  - pooled Poisson and random effects Poisson are inconsistent
  - need fixed effects explained below
  - or quasi first differences

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
```

variable name	storage type	display format	value 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

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

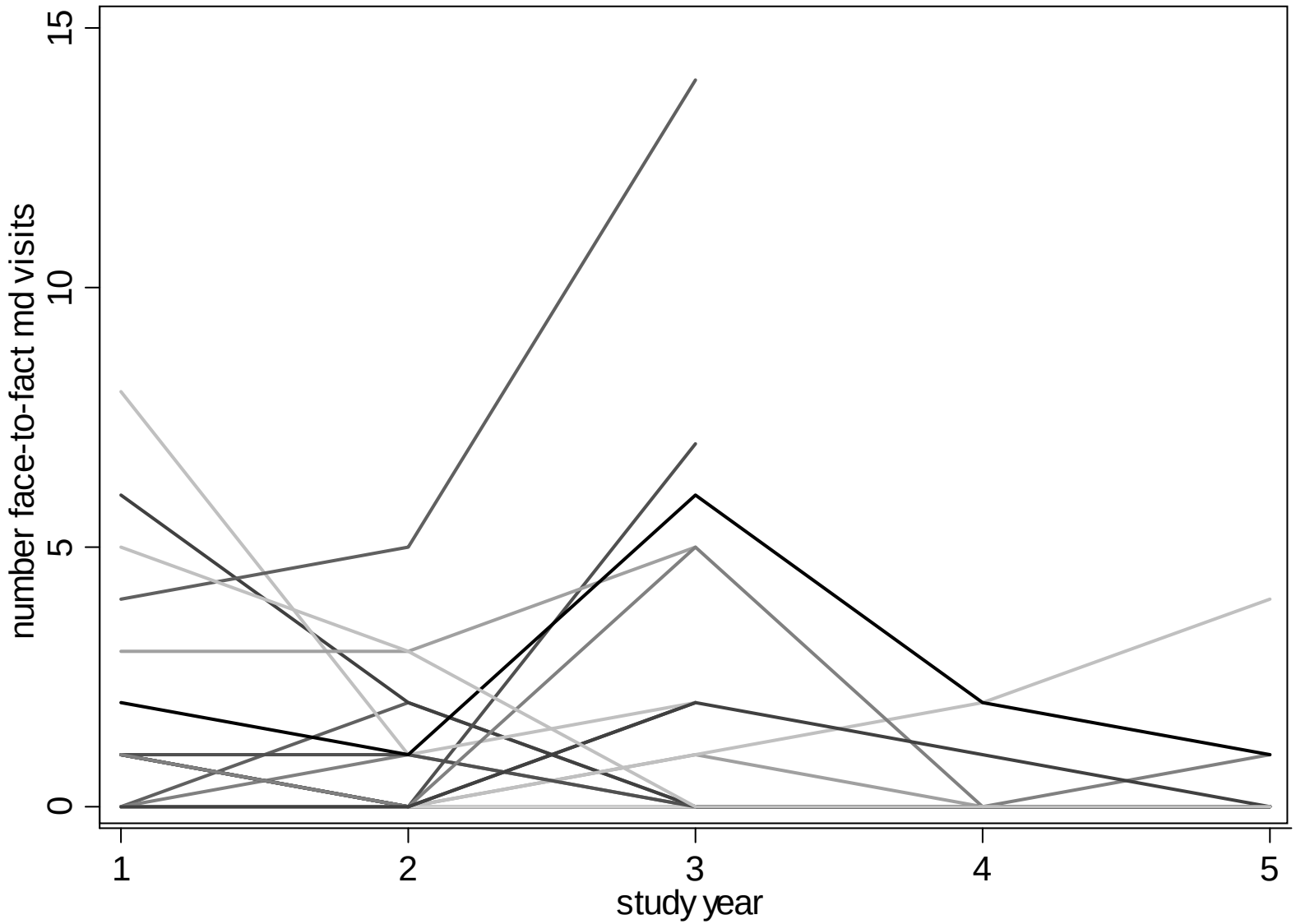
```
. summarize mdu lcoins ndisease female age lfam child id year
```

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

- Declare as panel data
  - `xtset id year`
- Panel is unbalanced. Most are in for 3 years or 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
```

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



- For mdu both within and between variation are important.

```
. * 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

- Only time-varying regressors are age, lfam and child
  - and these have mainly between variation
  - this will make within or fixed estimator very imprecise.

Panel Poisson

- Consider four panel Poisson estimators
  - Pooled Poisson with cluster-robust errors
  - Population-averaged Poisson (GEE)
  - Poisson random effects (gamma and normal)
  - Poisson fixed effects
- Can additionally apply most of these to negative binomial.
- And can extend FE to dynamic panel Poisson where $y_{i,t-1}$ is a regressor.

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)
------------------------------------------------------------------------------
             |               Robust
         mdu |      Coef.   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 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

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

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

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

- 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 $\mathsf{Cov}[y_{it}, y_{is}|\mathbf{x}_{it}, \mathbf{x}_{is}] \neq \rho$.

- 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, because do not control for overdispersion.

- The correlations $\text{Cor}[y_{it}, y_{is}|\mathbf{x}_i]$ 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
```

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 α_i is unobserved but is not correlated with $\mathbf{x}_{it}$.
- RE estimator 1: Assume α_i is $Gamma[1, \eta]$ distributed
  - closed-form solution exists (negative binomial)
  - $E[y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}] = \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$
- 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)
  - $E[y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}] \neq \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$!!! Now different conditional mean.

- Poisson random effects (gamma) with panel bootstrap 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)
------------------------------------------------------------------------------
             |               Robust
         mdu |      Coef.   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.

Panel Poisson method 4: 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 α_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 $N \longrightarrow \infty$.
  - but no problem as can eliminate α_i.
- Eliminate α_i by quasi-differencing as follows

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

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

- The final result implies

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

- Poisson fixed effects estimator solves the corresponding sample moment conditions

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

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

- Consistency requires

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

- 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)
------------------------------------------------------------------------------
             |               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
------------------------------------------------------------------------------
```

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 α_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 α_i

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

and α_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)
```

Comparison of 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
  - simplest to work with Poisson but make sure get cluster-robust standard errors to control for overdispersion.

Dynamic panels

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

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

  - Obvious model is $\alpha_i\lambda_{it} = \alpha_i \exp(\rho y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta})$ but this is potentially explosive.
  - May be better to use linear feedback: $\alpha_i\lambda_{it} = \alpha_i\{\rho y_{i,t-1} + \exp(\mathbf{x}_{it}'\boldsymbol{\beta})\}$
- Then $\mathsf{E}[y_{it} - (\lambda_{i,t-1}/\lambda_{it})y_{it-1}|y_{it-1},...,y_{i1},\mathbf{x}_{it},\ldots,\mathbf{x}_{i1}] = 0$
- So can do GMM based on moment condition

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

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

References

- In addition to those given in preceding set of slides.
- Finite Mixture Models
  - Deb, P. and P.K. Trivedi (2022), "The structure of demand for health care: latent class versus two-part models, " *Journal of health economics*, 2002.
- Panel Data
  - Cameron, A.C., and P.K. Trivedi (2015), "Count Panel Data," in B. Baltagi ed., *Oxford Handbook of Panel Data, Oxford: Oxford University Press*, pp.233-256.