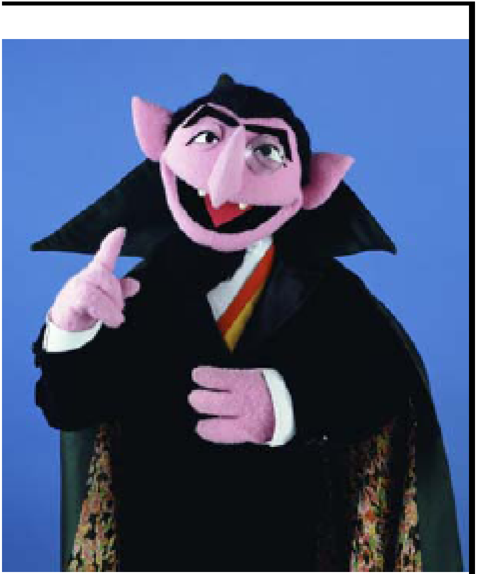
Recent Developments in Panel Models for Count Data

Pravin K. Trivedi
Indiana University. - Bloomington

Prepared for 2010 Mexican Stata Users Group meeting,
based on
A. Colin Cameron and Pravin K. Trivedi (2005),
Microeconometrics: Methods and Applications (**MMA**), C.U.P.
MMA, chapters 21-23
and
A. Colin Cameron and Pravin K. Trivedi (2010),
Microeconometrics using Stata Revised edition (**MUSR**), Stata Press.
MUSR, chapters 8;18.

April 29, 2010

0. Dedication



1. Introduction

- Objective 1: To survey recent developments in count data panel models
- Objective 2: Evaluate the advances made against background of main features of count data
- Objective 3: Highlight the areas where significant gaps exist and review the most promising approaches

Background (1)

- Panel data are repeated measures on individuals (i) over time (t): data are (y_{it}, $\mathbf{x}_{it}$) for $i = 1, ..., N$ and $t = 1, ..., T$, and y_{it} are nonnegative integer-valued outcomes.
- Conditional on $\mathbf{x}_{it}$, the y_{it} are likely to be serially correlated for a given i, partly because of state dependence and partly because of cserial correlation in shocks.
- Hence each additional year of data is not independent of previous years.
- Cross-sectional dependence between observations is also to be expected given emphasis on stratified clustered sampling designs.
- (1) Pervasive unobserved heterogeneity, (2) a typically high proportion of zeros, (3) inherent discreteness and heteroskedasticity generate complications that are hard to handle simultaneously
- Finally, the researcher's interest often goes beyond the conditional mean.
- How well does available software (Stata) handle these issues?

2. Basic linear panel models review

- Pooled model (or population-averaged)

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}. \quad (1)$$

- Two-way effects model allows intercept to vary over i and t

$$y_{it} = \alpha_i + \gamma_t + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}. \quad (2)$$

- Individual-specific effects model

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \quad (3)$$

  where α_i may be fixed effect or random effect.

- Mixed model or random coefficients model allows slopes to vary over i

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}. \quad (4)$$

3. Fixed effects versus random effects model

- Individual-specific effects model:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + (\alpha_i + \varepsilon_{it}).$$

- Fixed effects (FE):
  - α_i is a random variable possibly correlated with $\mathbf{x}_{it}$
  - so regressor $\mathbf{x}_{it}$ may be endogenous (wrt to α_i but not ε_{it})
    e.g. education is correlated with time-invariant ability
  - pooled OLS, pooled GLS, RE are inconsistent for $\boldsymbol{\beta}$
  - within (FE) and first difference estimators are consistent.
- Random effects (RE) or population-averaged (PA):
  - α_i is purely random (usually iid $(0, \sigma_\alpha^2)$) unrelated to $\mathbf{x}_{it}$
  - so regressor $\mathbf{x}_{it}$ is exogenous
  - appropraite FE and RE estimators are consistent for $\boldsymbol{\beta}$
- Fundamental divide: microeconometricians FE versus others RE.

4. Some features of nonlinear panel models

- In contrast to linear models, solutions for nonlinear models tend to lack generality and are model-specific.
- Standard count models include: Poisson and negative binomial
- General approaches are similar to those for the 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.
  - Count models involve discreteness, nonlinearity and intrinsic heteroskedasticity.

Some Standard Cross-section Count Models

	$f(y)$	$f(y) = \Pr[Y = y]$	Mean; Variance
1	Poisson	$e^{-\mu}\mu^{y}/y!$	$\mu(\mathbf{x})$; $\mu(\mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\beta})$
2	NB1	As in NB2 below with α^{-1} replaced by $\alpha^{-1}\mu$	$\mu(\mathbf{x})$; $(1+\alpha)\ \mu(\mathbf{x})$
3	NB2	$\frac{\Gamma(\alpha^{-1}+y)}{\Gamma(\alpha^{-1})\Gamma(y+1)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\frac{1}{\alpha}}\left(\frac{\mu}{\mu+\alpha^{-1}}\right)^{y}$	$\mu(\mathbf{x})$; $(1+\alpha\mu(\mathbf{x}))\ \mu(\mathbf{x})$
4	Hurdle	$\begin{cases} f_1(0) & \text{if } y = 0, \\ \frac{1-f_1(0)}{1-f_2(0)} f_2(y) & \text{if } y \geq 1. \end{cases}$	$\Pr[y > 0 \mid \mathbf{x}]\mathrm{E}_{y>0}[y \mid y > 0,$
5	ZI	$\begin{cases} f_1(0) + (1-f_1(0))f_2(0) & \text{if } y = 0, \\ (1-f_1(0))f_2(y) & \text{if } y \geq 1. \end{cases}$	$(1-f_1(0))\times$ $(\mu(\mathbf{x})+f_1(0)\mu^2(\mathbf{x}))$
6	FMM	$\sum_{j=1}^{m} \pi_j f_j(y \mid \boldsymbol{\theta}_j)$	$\Sigma_{i=1}^{2}\pi_i\mu_i(\mathbf{x})$; $\Sigma_{i=1}^{2}\pi_i[\mu_i(\mathbf{x})+\mu_i^2(\mathbf{x})]$

- A pooled or population-averaged (PA) model may be used.
  - This is same model as in cross-section case, with adjustment for correlation over time for a given individual.
- A fully parametric model may be specified, with **separable heterogeneity** and conditional density

$$f(y_{it}|\alpha_i, \mathbf{x}_{it}) = f(y_{it}, \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}, \gamma), \quad t = 1, ..., T_i, \; i = 1, ..., N, \tag{5}$$

  or **nonseparable heterogeneity**

$$f(y_{it}|\alpha_i, \mathbf{x}_{it}) = f(y_{it}, \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}_i, \gamma), \quad t = 1, ..., T_i, \; i = 1, ..., N, \tag{6}$$

  where γ denotes additional model parameters such as variance parameters and α_i is an individual effect.
- A semiparametric conditional mean (usually exponential mean) model may be specified, with additive effects

$$\mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}] = \alpha_i + g(\mathbf{x}_{it}'\boldsymbol{\beta}) \tag{7}$$

  or multiplicative effects

$$\mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}] = \alpha_i \times g(\mathbf{x}_{it}'\boldsymbol{\beta}). \tag{8}$$

5. Evolution of Panel Models (1)

- Focus on panel methods most commonly used by microeconometricians. The underlying asymptotic theory assumes short panels (T small, N large): data on many individual units and few time periods.

- The key paper in the modern treatment of panel analysis for counts is Hausman et al. (1984).
- The developments since 1984 can be summarized in generational terms as follows.

Evolution of Panel Models (2)

	G-1 models	G-2 models	G-3 models
Period	1974-1990	1991-2000	Post-2000
Function	Mainly parametric	Flexible parametric	Parametric / SP
CS models	Poisson, Negbin	Hurdles, finite mixtures, ZIP	Quantile reg; Selection models
Panel models	Poisson, Negbin	Poisson, Negbin, EM	EM; QR;
Unobs. hetero.	Multiplicative	Separable or nonsep.	Flexible; non-sep
Modeling α_i	Mainly RE or PA	RE, PA and fixed effects	RE/PA/FE/ Correlated RE; DV
Variance est.	Robust wrt overdispersion	Robust wrt OD	Robust or Cl-Rob wrt OD/SC
Dynamics	Lagged x's	Exponential feedback	Linear or exponential
Endogeneity	Largely ignored	Allowed in RE models	Allowed in RE and FE
Estimators	Mainly MLE	MLE; GEE; NLIV;	MLE; GEE; NLGMM; QR; QRIV

6. Remarks on the evolution of count panel models (2)

- FE panel data counterparts of several popular cross-section models like hurdles, FMM, and ZIP are undeveloped.
- When several complications occur simultaneously (e.g. nonseparable individual-specific effects and endogenous regressors) they are most conveniently analyzed in a RE or PA or moment-based models.
- Fully parametric methods for simultaneously handling endogeneity plus something else (e.g. nonseparable UH) are largely absent, and moment-based methods are a dominant alternative.
- Overdispersion-robust and cluster-robust estimation of variances is now feasible and very common.

7. Nonlinear: Pooled or population-averaged estimators

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

    `poisson y x, vce(cluster id)`

    or

    `xtgee y x, fam(poisson) link(log) corr(ind) vce(cluster id)`
- Extend pooled feasible GLS to the nonlinear case
  - Estimate with an assumed correlation structure over time
  - Equicorrelated probit example:

    `xtpoisson y x, pa vce(boot)`

    or

    `xtgee y x, fam(poisson) link(log) corr(exch) vce(cluster id)`

Nonlinear random effects estimators

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

$$\begin{aligned} & f(y_{it}, \ldots, y_{iT}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \boldsymbol{\beta}, \gamma, \boldsymbol{\eta}) \\ = & \int \left[\prod_{t=1}^{T} f(y_{it}|\mathbf{x}_{it}, \alpha_i, \boldsymbol{\beta}, \gamma)\right] g(\alpha_i|\boldsymbol{\eta}) d\alpha_i. \end{aligned} \tag{9}$$

- 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`

9. Finite Mixture or Latent Class model

- Suppose the sample is generated from the following dgp:

$$f(y_{it}|\mathbf{x}_{it},\Theta) = \sum_{j=1}^{C-1} \pi_j f_j(y_{it}|\mathbf{x}_{it},\boldsymbol{\theta}_j) + \pi_C f_C(y_i|\mathbf{x}_{it},\boldsymbol{\theta}_C), \quad (10)$$

where $\sum_{j=1}^{C} \pi_j = 1$, $\pi_j \geqslant 0$ $(j = 1, ..., C)$. For identifiability, use labelling restriction $\pi_1 \geq \pi_2 \geq .... \geq \pi_C$, always satisfied by rearrangement, postestimation.

- This specification accommodates discrete nonseparable heterogeneity between latent classes.
- Long history in statistics; see McLachlan and Basford (1988). Earlier treatments emphasized univariate formulations; (Lindsey, 1995) emphasized identification and complexity. Special cases: Heckman and Singer (1984)
- Probability distribution $f(y_i|\widehat{\Theta};\widehat{C})$ that maximizes $\mathcal{L}(\boldsymbol{\pi},\Theta,C|\mathbf{y})$ is called the semiparametric maximum likelihood estimator

- $f(y_i|\theta_j)$ can itself be a flexiblefunctional form that accommodates within-class heterogeneity
- C can be chosen using the hypothesis testing approach or model comparison approach
- 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.
- Can be implemented using Stata's `fmm` command such as

```
fmm y $xlist1, vce(robust) components(3) mixtureof(poisson)
```

10. Quantile regression

- The q^{th} quantile regression estimator $\widehat{\boldsymbol{\beta}}_q$ minimizes over $\boldsymbol{\beta}_q$

$$Q(\boldsymbol{\beta}_q) = \sum_{i:y_i \geq \mathbf{x}'_{it}\boldsymbol{\beta}}^{N} q|y_{it} - \mathbf{x}'_{it}\boldsymbol{\beta}_q| + \sum_{i:y_i < \mathbf{x}'_{it}\boldsymbol{\beta}}^{N} (1-q)|y_{it} - \mathbf{x}'_{it}\boldsymbol{\beta}_q|, \quad 0 < q < 1$$

  - Example: median regression with $q = 0.5$.
- Continuation transform: For count y adapt standard methods for continuous y by:
  - Replace count y by continuous variable $z = y + u$ where $u \sim Uniform[0,1]$.: "jittering step"
  - Then reconvert predicted z-quantile to y-quantile using ceiling function.
  - Machado and Santos Silva (JASA, 2005).

Adapting to the exponential mean

- Conventional count models based on exponential conditional mean, $\exp(\mathbf{x}'\boldsymbol{\beta})$, rather than $\mathbf{x}'\boldsymbol{\beta}$.
- $Q_q(y|\mathbf{x})$ and $Q_q(z|\mathbf{x})$ denote the q^{th} quantiles of the conditional distributions of y and z, respectively. To allow for exponentiation, $Q_q(z|\mathbf{x})$ is specified to be

$$Q_q(z|\mathbf{x}) = q + \exp(\mathbf{x}'\boldsymbol{\beta}_q).$$

- The additional term q appears because $Q_q(z|\mathbf{x})$ bounded from below by q, due to jittering.
- Log transformation is applied so that $\ln(z - q)$ is modelled, with the adjustment if $z - q < 0$
- Transformation justified by the property that quantiles are equivariant to monotonic transformation

Implementation

- Post-estimation transformation of the z-quantiles back to y-quantiles uses the ceiling function, with

$$Q_q(y|\mathbf{x}) = \lceil Q_q(z|\mathbf{x}) - 1 \rceil$$

where the symbol $\lceil r \rceil$ in the right-hand side denotes the smallest integer greater than or equal to r.

- To reduce the effect jittering the model is estimated multiple times using independent draws from $\mathcal{U}(0,1)$ distribution, and estimated coefficients and confidence interval endpoints are averaged. Hence the estimates of the quantiles of y counts are based on $\widehat{Q}_q(y|\mathbf{x}) = \lceil Q_q(z|\mathbf{x}) - 1 \rceil = \lceil q + \exp(\mathbf{x}'\overline{\widehat{\boldsymbol{\beta}}}_q) - 1 \rceil$, where $\overline{\widehat{\boldsymbol{\beta}}}$ denotes the average over the jittered replications.
- Variance estimation usually based on computationally intensive bootstrap

- QCR method of Machado and Santos Silva can be implemented using Stata add-on command `qcount`, due to Miranda (2006). The command syntax is:

  `qcount` *depvar* [*indepvars*] [*if*] [*in*], `quantile`(*number*) [, *repetition*(#)

where `quantile(`*number*`)` specifies the quantile to be estimated and *repetition*(#)specifies the number of jittered samples to be used to calculate the parameters of the model, the default value being 1000.

- Panel models can be estimated treating the data as repeated cross sections, as in PA approach.
- Main attraction is the ability to study differences in marginal effects at different quantiles.
- The post-estimation command `qcount_mfx` computes marginal effects for the model, evaluated at the means of the regressors.

QCR Example - Winkelmann, JHE 2006

- Using an unbalanced sample (1995-1999) from GSOEP, Winkelmann analyzes the differential impact of healthcare reform on distribution of doctor visits across quantiles.

R. Winkelmann / Journal of Health Economics 25 (2006) 131–145

Table 4
Relative frequencies of estimated y_α-quantiles for treatment group before and after reform

	0	1	2	3	4	5	6	7	8
Before reform									
$\hat{Q}_y(0.25\|x)$	67.54	21.50	10.70	0.26	0	0	0	0	0
$\hat{Q}_y(0.50\|x)$	0.89	57.85	19.77	7.12	4.38	5.03	4.38	0.55	0.02
$\hat{Q}_y(0.75\|x)$	0.89	36.23	21.62	10.81	8.96	5.07	2.05	2.00	2.39
$\hat{Q}_y(0.90\|x)$	0.17	14.30	23.04	14.45	9.50	6.05	6.95	4.13	3.96
After reform									
$\hat{Q}_y(0.25\|x)$	72.77	18.95	8.26	0.02	0	0	0	0	0
$\hat{Q}_y(0.50\|x)$	3.62	59.39	19.29	4.88	5.80	4.94	2.04	0.05	0
$\hat{Q}_y(0.75\|x)$	3.62	41.07	18.32	11.79	7.50	2.62	2.26	2.52	3.28
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

QR and panel data: pros and cons

- Excess zeros can make identification of lower quantiles difficult.
- Can QR accommodate fixed and random effects?
- Interpretation of fixed effects in QR context is somewhat tenuous; see Koenker (2004).
- QR has been extended to accommodate censoring, endogenous regressors; see Chernozhukov et al (2009)
- QR has also been extended to handle lagged dependent variable.

11. Nonlinear random slopes estimators

- Can extend to random slopes by adding an assumption about the distribution of 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); can be quite slow!
  - Developed by Sophia Rabe-Hesketh and Anders Skrondal.

12. Nonlinear fixed effects estimators

- In general not possible in short panels.
- Incidental parameters problem:
  - N fixed effects α_i plus K regressors means $(N+K)$ parameters
  - But $(N+K) \to \infty$ as $N \to \infty$
  - Need to eliminate α_i by some sort of differencing, or concentrated likelihood argument
  - possible for Poisson, negative binomial
- Stata commands
  - `xtpoisson, fe` (better to use `xtpqml` as robust se's)
  - `xtnbreg, fe`
- Fixed effects extensions to hurdle, finite mixture, zero-inflated models currently not available.

Incidental parameters in Poisson regression

- Derivation of fixed effects estimator for the Poisson panel
- Poisson MLE simultaneously estimates $\boldsymbol{\beta}$ and $\alpha_1, ..., \alpha_N$. The log-likelihood is

$$\begin{aligned}\ln L(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \ln \left[\prod_i \prod_t \{\exp(-\alpha_i \lambda_{it}) (\alpha_i \lambda_{it})^{y_{it}} / y_{it}!\}\right] \\ &= \sum_i \left[-\alpha_i \sum_t \lambda_{it} + \ln \alpha_i \sum_t y_{it} + \sum_t y_{it} \ln \lambda_{it} - \sum_t \ln y_{it}(\right.\end{aligned}$$

where $\lambda_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$.

- FOC with respect to α_i yields $\widehat{\alpha}_i = \sum_t y_{it} / \sum_t \lambda_{it}$ (a sufficient statistic for α_i)
- Substituting this yields the concentrated likelihood function.
- Dropping terms not involving $\boldsymbol{\beta}$,

$$\ln L_{\text{conc}}(\boldsymbol{\beta}) \propto \sum_i \sum_t \left[y_{it} \ln \lambda_{it} - y_{it} \ln \left(\sum_s \lambda_{is}\right)\right]. \qquad (12)$$

Interpretation

- Here is no incidental parameters problem.
- Consistent estimates of $\boldsymbol{\beta}$ for fixed T and $N \to \infty$ can be obtained by maximization of $\ln L_{\text{conc}}(\boldsymbol{\beta})$
- FOC with respect to $\boldsymbol{\beta}$ yields first-order conditions

$$\sum_i \sum_t \left[y_{it}\mathbf{x}_{it} - y_{it} \left[\sum_s \lambda_{is}\mathbf{x}_{is} \right] / \left[\sum_s \lambda_{is} \right] \right] = \mathbf{0},$$

that can be re-expressed as

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it} \left(y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i} \bar{y}_i \right) = \mathbf{0}, \tag{13}$$

FE Poisson: pros and cons

- Time-invariant regressors will be eliminated also by the transformation. Some marginal effects not identified.
- May substitute individual specific dummy variables, though this raises some computational issues.
- Poisson and linear panel model special in that simultaneous estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ provides consistent estimates of $\boldsymbol{\beta}$ in short panels, so there is no incidental parameters problem.
- The above assumes strict exogeneity of regressors.
- We can handle endogenous regressors under weak exogeneity assumption. A moment condition estimator can be defined using the FOC (13).
- This FE approach **does not extend** to several empiricaly important models: hurdle, fmm, and zip.

Ad hoc methods for handling fixed effects

- Are we making too much of the fixed effects and the associated incidental paramnetr problem?
- The dummy variables solution; Allison (2009); Greene (2004).

13. Stata Commands

- Nonlinear panel estimators

Estimator	Count data
Pooled	`poisson; nbreg`
Quantile	`qcount q(%), rep(#)`
FMM	`fmm components(#) mixtureof(poisson)` `fmm components(#) mixtureof(nbreg)`
GEE (PA)	xtgee,family(poisson)
	xtgee,family(nbinomial)
RE	`xtpoisson, re`
	`xtnbreg, fe`
Random slopes	`xtmepoisson`
FE	`xtpoisson, fe`
	`xtnbreg, fe`

- FMM is not part of official Stata but is in the public domain and can be added

Panel counts: data example

- 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
```

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

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
```

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.

14. 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.

Model	Moment spec.	Estimating equations
Pooled Poisson PA	$\mathrm{E}[y_{it}\|x_{it}] = \exp(x_{it}'\beta)$	$\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}(y_{it} - \mu_{it}) = 0$ where $\mu_{it} = \exp(x_{it}'\beta)$ $\rho_{ts} = \mathrm{Cor}[(y_{it} - \exp(x_{it}'\beta))(y_{is} - \exp$
RE Poisson	$\mathrm{E}[y_{it}\|\alpha_i, x_{it}] = \alpha_i \exp(x_{it}'\beta)$	$\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}\left(y_{it} - \mu_{it}\frac{\bar{y}_i + \eta/T}{\overline{\mu}_i + \eta/T}\right)$ $\overline{\mu}_i = T^{-1}\sum_t \exp(x_{it}'\beta);\ \eta = \mathrm{var}(\alpha_i)$
FE Pois	$\mathrm{E}[y_{it}\|\alpha_i, x_{it}] = \alpha_i \exp(x_{it}'\beta)$	$\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}\left(y_{it} - \mu_{it}\frac{\bar{y}_i}{\overline{\mu}_i}\right) = 0,$

15. 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!!

16. Panel Poisson method 2: population-averaged

- Assume that for the i^{th} observation moments are like for GLM Poisson

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

- Stack the conditional means for the i^{th} individual:

$$\mathsf{E}[\mathbf{y}_i|\mathbf{X}_i] = \mathbf{m}_i(\boldsymbol{\beta}) = \begin{bmatrix} \exp(\mathbf{x}_{i1}'\boldsymbol{\beta}) \\ \vdots \\ \exp(\mathbf{x}_{iT}'\boldsymbol{\beta}) \end{bmatrix}.$$

  where $\mathbf{y}_i = [\mathbf{y}_{i1}, ..., \mathbf{y}_{iT}]'$ and $\mathbf{X}_i = [\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}]'$.
- Stack the conditional variances for the i^{th} individual.
  - With no correlation

$$\mathsf{V}[y_i|\mathbf{X}_i] = \phi\mathbf{H}_i(\boldsymbol{\beta}) = \phi \times \mathrm{Diag}[\exp(\mathbf{x}_{it}'\boldsymbol{\beta})].$$

- Assume a pattern $\mathbf{R}(\boldsymbol{\rho})$ for autocorrelation over t for given i so

$$\mathsf{V}[y_i|\mathbf{X}_i] = \phi\mathbf{H}_i(\boldsymbol{\beta})^{1/2}\mathbf{R}(\boldsymbol{\rho})\mathbf{H}_i(\boldsymbol{\beta})^{1/2}$$

- This is called a working matrix.
  - Example: $\mathbf{R}(\boldsymbol{\rho}) = \mathbf{I}$ if there is no correlation
  - Example: $\mathbf{R}(\boldsymbol{\rho}) = \mathbf{R}(\rho)$ has diagonal entries 1 and off diagonal entries ρ if there is equicorrelation.
  - Example: $\mathbf{R}(\boldsymbol{\rho}) = \mathbf{R}$ where diagonal entries 1 and off-diagonals unrestricted (< 1).

17. Stata's GEE command

- The GLM estimator solves: $\sum_{i=1}^{N} \frac{\partial \mathbf{m}_i'(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \mathbf{H}_i(\boldsymbol{\beta})^{-1}(\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\theta})) = \mathbf{0}$.
- Generalized estimating equations (GEE) estimator or population-averaged estimator (PA) of Liang and Zeger (1986) solves

$$\sum_{i=1}^{N} \frac{\partial \mathbf{m}_i'(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \widehat{\Omega}_i^{-1}(\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\beta})) = \mathbf{0},$$

where $\widehat{\Omega}_i$ equals Ω_i in with $\mathbf{R}(\boldsymbol{\alpha})$ replaced by $\mathbf{R}(\widehat{\boldsymbol{\alpha}})$ where $\text{plim}\, \widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$.

- Cluster-robust estimate of the variance matrix of the GEE estimator is

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_{\mathrm{GEE}}] = \left(\widehat{\mathbf{D}}'\widehat{\Omega}^{-1}\widehat{\mathbf{D}}\right)^{-1} \left(\sum_{g=1}^{G} \mathbf{D}_g'\widehat{\Omega}_g^{-1}\widehat{\mathbf{u}}_g\widehat{\mathbf{u}}_g'\widehat{\Omega}_g^{-1}\mathbf{D}_g\right) \left(\mathbf{D}'\widehat{\Omega}^{-1}\mathbf{D}\right)^{-1},$$

where $\widehat{\mathbf{D}}_g = \partial \mathbf{m}_g'(\boldsymbol{\beta})/\partial \boldsymbol{\beta}\big|_{\widehat{\boldsymbol{\beta}}}$, $\widehat{\mathbf{D}} = [\widehat{\mathbf{D}}_1, ..., \widehat{\mathbf{D}}_G]'$, $\widehat{\mathbf{u}}_g = \mathbf{y}_g - \mathbf{m}_g(\widehat{\boldsymbol{\beta}})$, and now $\widehat{\Omega}_g = \mathbf{H}_g(\widehat{\boldsymbol{\beta}})^{1/2}\mathbf{R}(\widehat{\boldsymbol{\rho}})\mathbf{H}_g(\widehat{\boldsymbol{\beta}})^{1/2}$.

  - The asymptotic theory requires that $G \to \infty$.

Population-averaged Poisson with unstructured correlation

```
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,
No 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
```

18. Panel Poisson method 3: random effects

- 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)
  - $\mathrm{E}[y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}] = \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$
- 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)
  - $\mathrm{E}[y_{it}|\mathbf{x}_{it}, \boldsymbol{\beta}] = \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$ aside from translation of intercept.

Poisson random effects (gamma) with panel bootstrap se's

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

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

                                                Wald chi2(6)       =    529.10
Log likelihood  = -43240.556                    Prob > chi2        =    0.0000

                                   (Replications based on 5908 clusters in id)
```

mdu	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
lcoins	-.0878258	.0086097	-10.20	0.000	-.1047004	-.0709511
ndisease	.0387629	.0026904	14.41	0.000	.0334899	.0440359
female	.1667192	.0379216	4.40	0.000	.0923942	.2410442
age	.0019159	.0016242	1.18	0.238	-.0012675	.0050994
lfam	-.1351786	.0308529	-4.38	0.000	-.1956492	-.0747079
child	.1082678	.0495487	2.19	0.029	.0111541	.2053816
_cons	.7574177	.0754536	10.04	0.000	.6095314	.905304
/lnalpha	.0251256	.0270297			-.0278516	.0781029
alpha	1.025444	.0277175			.9725326	1.081234

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

19. Poisson fixed effects with panel bootstrap se's

```
. xtpoisson mdu lcoins ndisease female age lfam child, fe vce(boot, reps(100) seed(10
(running xtpoisson on estimation sample)

Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

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)       =      4.64
Log likelihood  = -24173.211                    Prob > chi2        =    0.2002

                                 (Replications based on 4977 clusters in id)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0112009   .0095077    -1.18   0.239    -.0298356    .0074339
        lfam |   .0877134   .1125783     0.78   0.436     -.132936    .3083627
       child |   .1059867   .0738452     1.44   0.151    -.0387472    .2507206
------------------------------------------------------------------------------
```

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

- 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

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

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 RE models
  - note that these data are not good to illustrate FE as regressors have little within variation.

20. Comparison of different Poisson estimators with cluster-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	0.0086	0.0073	
ndisease	0.0339	0.0346	0.0388	0.0409	
	0.0026	0.0024	0.0027	0.0023	
female	0.1718	0.1585	0.1667	0.2084	
	0.0343	0.0334	0.0379	0.0305	
age	0.0041	0.0031	0.0019	0.0027	-0.0112
	0.0017	0.0015	0.0016	0.0012	0.0095
lfam	-0.1482	-0.1407	-0.1352	-0.1443	0.0877
	0.0323	0.0294	0.0309	0.0265	0.1126
child	0.1030	0.1014	0.1083	0.0737	0.1060
	0.0507	0.0430	0.0495	0.0345	0.0738
_cons	0.7488	0.7765	0.7574	0.2873	
	0.0786	0.0717	0.0755	0.0642	
lnalpha					
_cons			0.0251		
			0.0270		
lnsig2u					
_cons				0.0550	
				0.0255	
Statistics					

- Predetrmined means regressor correlated with current and past shoocks but not future shocks: $E[u_{it}x_{is}] = 0$ for $s \geq t$, but $\neq 0$ for $S < t$.

- Two specifications are considered:

$$\begin{aligned} y_{it} &= \exp(\mathbf{x}_{it}'\beta)\nu_i w_{it} \\ y_{it} &= \exp(\mathbf{x}_{it}'\beta)\nu_i + w_{it} \end{aligned}$$

- A quasi-differencing transformation is used to eliminate the fixed effect.

- Then a moment condition is constructed for estimation.

- Depending upon which specification is used different moment conditions obtain.

- Chamberlain and Wooldridge derive quasi-differencing transformations that are shown in the table below.

21. Exponential Mean and Multiplicative Heterogeneity

- Relies on a number of ways of eliminating the fixed effects
- Error may enter additively or multiplicatively
- Estimating equations are orthogonality conditions after quasi-differencing which eliminates the fixed effect

Model	**Moment spec.**	**Estimating equations**
Strict exog.	$\mathrm{E}[\mathbf{x}_{it} u_{it+j}] = 0,\ j \geq 0$	
Predetermined regressors	$\mathrm{E}[\mathbf{x}_{it} u_{it-s}] \neq 0,\ s \geq 1$	
GMM	Chamberlain	$\mathrm{E}\left[y_{it} \dfrac{\lambda_{it-1}}{\lambda_{it}} - y_{it-1} \mid \mathbf{x}_i^{t-1}) \right] = 0$
	Wooldridge	$\mathrm{E}\left[\dfrac{y_{it}}{\lambda_{it}} - \dfrac{y_{it-1}}{\lambda_{it-1}} \mid \mathbf{x}_i^{t-1}) \right] = 0$
GMM/endog	Wooldridge	$\mathrm{E}\left[\dfrac{y_{it}}{\lambda_{it}} - \dfrac{y_{it-1}}{\lambda_{it-1}} \mid \mathbf{x}_i^{t-2}) \right] = 0$

1. Use an interactive version of an estimation command (e.g. gmm); enter the function directly on the command line or dialog box by using a *substitutable expression*.
2. Use a *function evaluator program* which gives more flexibility in defining your objective function; usually more complicated to use but may be needed for more complicated problems.

- Hint: In Stata a good place to start is the nl (nonlinear least squares) command. Then go on to gmm.
- Most of the examples here involve substitutable expressions. Examples of function evaluator programs are in MUS and especially in Stata manuals.
- Example: $\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}\left(y_{it} - \mu_{it}\frac{\bar{y}_i}{\bar{\mu}_i}\right) = 0,$

22. Applications using balanced panel MEPS data

- For illustrating panel methods the RAND data set has limitations

```
. sum  officevis TOofficevis educ age income totchr
```

Variable	Obs	Mean	Std. Dev.	Min	Max
officevis	78888	1.387372	3.328148	0	94
TOofficevis	78888	1.488084	3.334559	0	58
educ	78888	12.32776	3.264869	0	17
age	78888	4.562129	1.742034	1.8	8.5
income	78888	27.60833	28.94855	-63.631	264.674
totchr	78888	.7881047	1.081315	0	7

MEPS Data

- Quarterly data for 2005-06

```
. xtdes

dupersid:  30002019, 30004010, ..., 38505016                  n =       9861
timeindex:  1, 2, ..., 8                                      T =          8
           Delta(timeindex) = 1 unit
           Span(timeindex)  = 8 periods
           (dupersid*timeindex uniquely identifies each observation)

Distribution of T_i:   min      5%      25%       50%       75%      95%      max
                         8       8        8         8         8        8        8

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+----------
     9861    100.00  100.00 |  11111111
 ---------------------------+----------
     9861    100.00         |  XXXXXXXX
```

Fixed Effects GMM in Stata 11

```
. program gmm_poi2
      version 11
      syntax varlist if, at(name) myrhs(varlist) ///
>    mylhs(varlist) myidvar(varlist)
      quietly {
      tempvar mu mubar ybar
      gen double `mu' = 0 `if'
      local j = 1
      foreach var of varlist `myrhs' {
         replace `mu' = `mu' + `var'*`at'[1,`j'] `if'
         local j = `j' + 1
         }
      replace `mu' = exp(`mu')
      egen double `mubar' = mean(`mu') `if', by(`myidvar')
      egen double `ybar' = mean(`mylhs') `if', by(`myidvar')
      replace `varlist' = `mylhs' - `mu'*`ybar'/`mubar' `if'
      }
  end
```

Implementing fixed effects GMM in Stata 11

```
. gmm gmm_poi2, mylhs(officevis) myrhs(insprv age income totchr)   ///
>  myidvar(dupersid) nequations(1) parameters(insprv age income totchr)   ///
>  instruments(insprv age income totchr, noconstant) onestep

Step 1
Iteration 0:   GMM criterion Q(b) =  .00140916
Iteration 1:   GMM criterion Q(b) =  1.487e-07
Iteration 2:   GMM criterion Q(b) =  1.583e-14
Iteration 3:   GMM criterion Q(b) =  1.843e-28

GMM estimation

Number of parameters =   4
Number of moments    =   4
Initial weight matrix: Unadjusted                     Number of obs   =   78888

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     /insprv |  -.0080549   .5460749    -0.01   0.988    -1.078342    1.062232
        /age |  -.5125841    13.1682    -0.04   0.969    -26.32178    25.29662
     /income |    .001128   .0013911     0.81   0.417    -.0015984    .0038545
     /totchr |   .2211125   .3354182     0.66   0.510    -.4362951    .8785201
------------------------------------------------------------------------------
Instruments for equation 1: insprv age income totchr

. estimates store PFEGMM
```

Standard fixed effects panel Poisson

```
. * Usual panel Poisson FE
. xtpoisson officevis insprv age income totchr, fe
note: 1900 groups (15200 obs) dropped because of all zero outcomes

Iteration 0:   log likelihood = -84468.435
Iteration 1:   log likelihood =  -84154.68
Iteration 2:   log likelihood = -84154.647
Iteration 3:   log likelihood = -84154.647

Conditional fixed-effects Poisson regression    Number of obs      =     63688
Group variable: dupersid                        Number of groups   =      7961

                                                Obs per group: min =         8
                                                               avg =       8.0
                                                               max =         8

                                                Wald chi2(4)       =    618.20
Log likelihood  = -84154.647                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
   officevis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      insprv |  -.0080549    .027985    -0.29   0.773    -.0629046    .0467947
         age |  -.5125841   .0629145    -8.15   0.000    -.6358943   -.3892739
      income |    .001128    .000258     4.37   0.000     .0006224    .0016336
      totchr |   .2211125   .0091051    24.28   0.000     .2032669    .2389582
------------------------------------------------------------------------------

. estimates store PFE
```

Standard FE Poisson with robust SE (with xtpqml add-on)

```
. * Add-on xtpqml gives panel robust se's
. xtpqml officevis insprv age income totchr, fe i(dupersid)
note: 1900 groups (15200 obs) dropped because of all zero outcomes

Iteration 0:   log likelihood = -84468.435
Iteration 1:   log likelihood =  -84154.68
Iteration 2:   log likelihood = -84154.647
Iteration 3:   log likelihood = -84154.647

Conditional fixed-effects Poisson regression    Number of obs      =     63688
Group variable: dupersid                        Number of groups   =      7961

                                                Obs per group: min =         8
                                                               avg =       8.0
                                                               max =         8

                                                Wald chi2(4)       =    618.20
Log likelihood  = -84154.647                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
   officevis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      insprv |  -.0080549    .027985    -0.29   0.773    -.0629046    .0467947
         age |  -.5125841   .0629145    -8.15   0.000    -.6358943   -.3892739
      income |    .001128    .000258     4.37   0.000     .0006224    .0016336
      totchr |   .2211125   .0091051    24.28   0.000     .2032669    .2389582
------------------------------------------------------------------------------
Calculating Robust Standard Errors...
------------------------------------------------------------------------------
   officevis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
officevis    |
      insprv |  -.0080549   .0715881    -0.11   0.910    -.1483651    .1322552
         age |  -.5125841   .1804831    -2.84   0.005    -.8663245   -.1588438
      income |    .001128   .0007661     1.47   0.141    -.0003734    .0026295
      totchr |   .2211125   .0250814     8.82   0.000     .1719539    .2702712
------------------------------------------------------------------------------
```

23. Panel dynamic

- Individual effects model allows for time series persistence via unobserved heterogeneity (α_i)
  - e.g. High α_i means high doctor visits each period
- Alternative time series persistence is via true state dependence (y_{t-1})
  - e.g. Many doctor visits last period lead to many this period.
- Linear model:

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

- Poisson model with exponnetial feedback: One possibility (designed to confront the zero problem) is

$$\begin{aligned} \mu_{it} &= \alpha_i \lambda_{it-1} = \alpha_i \exp(\rho y^*_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta}), \\ y^*_{i,t-1} &= \min(c, y_{i,t-1}). \end{aligned}$$

Panel dynamic: GMM estimation of FE model

- In fixed effects case Poisson FE estimator is now inconsistent.
- Instead assume **weak exogeneity**

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

- And use an alternative quasi-difference

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

- So MM or GMM based on

$$\mathrm{E}\left[\mathbf{z}_{it}\left(y_{it}-\frac{\lambda_{it-1}}{\lambda_{it}}y_{it-1}\right)\right]=\mathbf{0}$$

  where e.g. $\mathbf{z}_{it}=(y_{it-1},\mathbf{x}_{it})$ in just-identified case.
- Windmeijer (2008) has recent discussion.

Example of dynamic moment-based JI GMM

Ignore individual specific effects

```
. gmm (officevis - exp({xb:L.officevis insprv educ age income totchr}+{b0})),   ///
>  instruments(L.officevis insprv educ age income totchr) onestep vce(cluster dupersid)

Step 1
Iteration 0:   GMM criterion Q(b) =  4.9539327
Iteration 1:   GMM criterion Q(b) =  4.7296297
Iteration 2:   GMM criterion Q(b) =  1.4832673
Iteration 3:   GMM criterion Q(b) =  .01045573
Iteration 4:   GMM criterion Q(b) =  6.508e-06
Iteration 5:   GMM criterion Q(b) =  3.032e-12
Iteration 6:   GMM criterion Q(b) =  7.264e-25

GMM estimation

Number of parameters =   7
Number of moments    =   7
Initial weight matrix: Unadjusted                     Number of obs   =    69027

                           (Std. Err. adjusted for 9861 clusters in dupersid)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/xb_L_offi~s	.064072	.0041069	15.60	0.000	.0560228	.0721213
/xb_insprv	.2152153	.0331676	6.49	0.000	.1502079	.2802227
/xb_educ	.0404143	.0065808	6.14	0.000	.0275162	.0533124
/xb_age	.1221278	.0134542	9.08	0.000	.0957581	.1484976
/xb_income	-.0003585	.0004981	-0.72	0.472	-.0013347	.0006178
/xb_totchr	.3027348	.0141805	21.35	0.000	.2749415	.330528
/b0	-1.447292	.0952543	-15.19	0.000	-1.633987	-1.260597

Example of dynamic moment-based OI GMM

```
. gmm (officevis - exp({xb:L.officevis insprv educ age income totchr}+{b0})),   ///
>  instruments(L.officevis educ age income totchr female white hispanic married employed) /
>  onestep vce(cluster dupersid)

Step 1
Iteration 0:   GMM criterion Q(b) =  4.9696148
Iteration 1:   GMM criterion Q(b) =  3.7545442
Iteration 2:   GMM criterion Q(b) =  .86353039
Iteration 3:   GMM criterion Q(b) =  .25844389
Iteration 4:   GMM criterion Q(b) =  .07248002
Iteration 5:   GMM criterion Q(b) =  .07235453
Iteration 6:   GMM criterion Q(b) =  .07235443

GMM estimation

Number of parameters =   7
Number of moments    =  11
Initial weight matrix: Unadjusted                     Number of obs   =    69027

                            (Std. Err. adjusted for 9861 clusters in dupersid)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/xb_L_offi~s	.0631186	.0042901	14.71	0.000	.0547101	.071527
/xb_insprv	.0468067	.1154105	0.41	0.685	-.1793937	.273007
/xb_educ	.0422612	.0074362	5.68	0.000	.0276866	.0568359
/xb_age	.1208516	.0136986	8.82	0.000	.0940028	.1477003
/xb_income	.0004412	.0007107	0.62	0.535	-.0009518	.0018341
/xb_totchr	.2988192	.0144326	20.70	0.000	.2705318	.3271066
/b0	-1.361726	.0972536	-14.00	0.000	-1.55234	-1.171113

24. Poisson Extensions

- A different ML approach to dynamic specification

$$
\begin{aligned}
y_{i,t} &\sim P(\lambda_{it}),\ i = 1, ..., N; t = 1, ..., T \\
f(y_{i,t}|\lambda_{it}) &= \frac{e^{-\lambda_{it}}\lambda_{it}^{y_{it}}}{y_{it}!} \\
\lambda_{it} &= \nu_{it}\mu_{it} = \mathrm{E}[y_{it}|y_{i,t-1}, \mathbf{x}_{it}, \alpha_i] = g(y_{i,t-1}, \mathbf{x}_{it}, \alpha_i)
\end{aligned}
$$

- Initial conditions problem in dynamic model. In a short panel bias induced by neglect of dependence on initial condition.
- The lagged dependent variable on the right hand side a source of bias because the lagged dependent variable and individual-specific effect are correlated.
- Wooldridge's method (2005) integrates out the individual-specific random effect after conditioning on the initial value and covariates. Random effect model used to accommodate the initial conditions.

Alternative specifications

$$\mathsf{E}[y_{it}|\mathbf{x}_{it}, y_{it-1}, \alpha_i] = h(y_{it}, \mathbf{x}_{it}, \boldsymbol{\alpha}_i)$$

where α_i is the individual-specific effect.

- 1st alternative: Autoregressive dependence through the exponential mean.

$$\mathsf{E}[y_{it}|\mathbf{x}_{it}, y_{it-1}, \alpha_i] = \exp(\rho y_{it-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\alpha}_i)$$

- If the α_i are uncorrelated with the regressors, and further if parametric assumptions are to be avoided, then this model can be estimated using either the nonlinear least squares or pooled Poisson MLE. In either case it is desirable to use the robust variance formula.
- Limitation: Potentially explosive if large values of y_{it} are realized.

Initial conditions

- Dynamic panel model requires additional assumptions about the relationship between the initial observations ("initial conditions") $\mathbf{y}_0$ and the α_i.
- Effect of initial value on the future events is important in a short panel. The initial-value effect might be a part of individual-specific effect
- Wooldridge's method requires a specification of the conditional distribution of α_i given $\mathbf{y}_0$ and $\mathbf{z}_i$, with the latter entering separably.
- Under the assumption that the initial conditions are nonrandom, the standard random effects conditional maximum likelihood approach identifies the parameters of interest.
- For a class of nonlinear dynamic panel models, including the Poisson model, Wooldridge (2005) analyzes this model which **conditions** the joint distribution on the initial conditions.

Conditionally correlated RE (1)

- **Where parametric FE models are not feasible, the conditionally correlated random (CCR) effects model** (Mundlak (1978) and Chamberlain (1984)) **provides a compromise between FE and RE models.**
- Standard RE panel model assumes that α_i and $\mathbf{x}_{it}$ are uncorrelated. Making α_i a function of $\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}$ allows for possible correlation:

$$\alpha_i = \mathbf{z}_i' \boldsymbol{\lambda} + \varepsilon_i$$

- Mundlak's (more parsimonious) method allows the individual-specific effect to be determined by time averages of covariates, denoted $\mathbf{z}_i$; Chamberlain's method suggests a richer model with a weighted sum of the covariates for the random effect.

Conditionally correlated RE (2)

- We can further allow for initial condition effect by including $\mathbf{y}_0$ thus:

$$\alpha_i = \mathbf{y}_0' \boldsymbol{\eta} + \mathbf{z}_i' \boldsymbol{\lambda} + \varepsilon_i$$

  where $\mathbf{y}_0$ is a vector of initial conditions, $\mathbf{z}_i = \bar{\mathbf{x}}_i$ denotes the time-average of the exogenous variables and ε_i may be interpreted as unobserved heterogeneity.
- The formulation essentially introduces no additional problems though the averages change when new data are added. Estimation and inference in the pooled Poisson or NLS model can proceed as before.
- Formulation can also be used when no dynamics are present in the model. In this case ε_i can be integrated out using a distributional assumption about $f(\varepsilon)$.

Dynamic GMM without initial condition

- Here individual specific effect is captured by the initial condition

```
. gmm (officevis - exp({xb:L.officevis insprv educ age income totchr}+{b0})),   ///
>  instruments(L.officevis insprv educ age income totchr) onestep vce(cluster dupersid)

Step 1
Iteration 0:   GMM criterion Q(b) =  4.9539327
Iteration 1:   GMM criterion Q(b) =  4.7296297
Iteration 2:   GMM criterion Q(b) =  1.4832673
Iteration 3:   GMM criterion Q(b) =  .01045573
Iteration 4:   GMM criterion Q(b) =  6.508e-06
Iteration 5:   GMM criterion Q(b) =  3.032e-12
Iteration 6:   GMM criterion Q(b) =  7.264e-25

GMM estimation

Number of parameters =   7
Number of moments    =   7
Initial weight matrix: Unadjusted                     Number of obs   =    69027

                              (Std. Err. adjusted for 9861 clusters in dupersid)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/xb_L_offi~s	.064072	.0041069	15.60	0.000	.0560228	.0721213
/xb_insprv	.2152153	.0331676	6.49	0.000	.1502079	.2802227
/xb_educ	.0404143	.0065808	6.14	0.000	.0275162	.0533124
/xb_age	.1221278	.0134542	9.08	0.000	.0957581	.1484976
/xb_income	-.0003585	.0004981	-0.72	0.472	-.0013347	.0006178
/xb_totchr	.3027348	.0141805	21.35	0.000	.2749415	.330528
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Overidentified dynamic GMM with initial condition

```
. gmm (officevis - exp({xb:L.officevis insprv educ age income totchr}+{b0})),   ///
>  instruments(L.officevis educ age income totchr female white hispanic married empl
>  onestep vce(cluster dupersid)

Step 1
Iteration 0:   GMM criterion Q(b) =  4.9696148
Iteration 1:   GMM criterion Q(b) =  3.7545442
Iteration 2:   GMM criterion Q(b) =  .86353039
Iteration 3:   GMM criterion Q(b) =  .25844389
Iteration 4:   GMM criterion Q(b) =  .07248002
Iteration 5:   GMM criterion Q(b) =  .07235453
Iteration 6:   GMM criterion Q(b) =  .07235443

GMM estimation

Number of parameters =   7
Number of moments    =  11
Initial weight matrix: Unadjusted                     Number of obs   =   69027

                           (Std. Err. adjusted for 9861 clusters in dupersid)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
/xb_L_offi~s |   .0631186   .0042901    14.71   0.000     .0547101     .071527
  /xb_insprv |   .0468067   .1154105     0.41   0.685    -.1793937     .273007
    /xb_educ |   .0422612   .0074362     5.68   0.000     .0276866    .0568359
     /xb_age |   .1208516   .0136986     8.82   0.000     .0940028    .1477003
  /xb_income |   .0004412   .0007107     0.62   0.535    -.0009518    .0018341
  /xb_totchr |   .2988192   .0144326    20.70   0.000     .2705318    .3271066
```

Dynamic Just Identified GMM with Initial Conditions

```
. gmm (officevis - exp({xb:L.officevis T0officevis insprv educ age income totchr}+{b
> instruments(L.officevis T0officevis insprv educ age income totchr) onestep vce(cl

Final GMM criterion Q(b) =  6.30e-26

GMM estimation

Number of parameters =   8
Number of moments    =   8
Initial weight matrix: Unadjusted                     Number of obs  =   69027

                           (Std. Err. adjusted for 9861 clusters in dupersid)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
/xb_L_offi~s |   .0495929   .0044248    11.21   0.000     .0409204    .0582654
/xb_T0offi~s |   .0311947   .0043446     7.18   0.000     .0226794    .0397099
  /xb_insprv |   .2153361   .0351702     6.12   0.000     .1464038    .2842684
    /xb_educ |   .0382539   .0056386     6.78   0.000     .0272024    .0493054
     /xb_age |   .1303702   .0095834    13.60   0.000      .111587    .1491534
  /xb_income |  -.0003019   .0004701    -0.64   0.521    -.0012232    .0006194
  /xb_totchr |   .2847798    .010334    27.56   0.000     .2645256    .3050341
         /b0 |  -1.484486   .0605323   -24.52   0.000    -1.603127   -1.365845
------------------------------------------------------------------------------
Instruments for equation 1: L.officevis T0officevis insprv educ age income totchr _c
```

Dynamic Over Identified GMM with Initial Condition

```
. gmm (officevis - exp({xb:L.officevis T0officevis insprv educ age income totchr}+{b
> instruments(L.officevis T0officevis educ age income totchr female white hispanic
> onestep vce(cluster dupersid) nolog

Final GMM criterion Q(b) =  .0685762

GMM estimation

Number of parameters =   8
Number of moments    =  12
Initial weight matrix: Unadjusted                      Number of obs   =   69027

                          (Std. Err. adjusted for 9861 clusters in dupersid)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
/xb_L_offi~s |   .0490201   .0046062    10.64   0.000      .039992    .0580481
/xb_T0offi~s |   .0305356   .0044538     6.86   0.000     .0218063    .0392648
  /xb_insprv |   .0565968   .1135886     0.50   0.618    -.1660328    .2792264
    /xb_educ |   .0402952   .0059253     6.80   0.000     .0286819    .0519085
     /xb_age |   .1299791   .0098075    13.25   0.000     .1107567    .1492014
  /xb_income |   .0004368    .000703     0.62   0.534    -.0009411    .0018148
  /xb_totchr |   .2805608   .0101571    27.62   0.000     .2606532    .3004684
         /b0 |  -1.408679   .0607941   -23.17   0.000    -1.527833   -1.289525
------------------------------------------------------------------------------
Instruments for equation 1: L.officevis T0officevis educ age income totchr female wh
    married employed _cons
```

Alternative to EFM: LFM

- An alternative to the (potentially explosive) EF is the **linear feedback model**

$$\mathrm{E}[y_{it}|\mathbf{x}_{it}, y_{it-1}, \alpha_i] = \rho^* y_{it-1} + \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\alpha}_i)$$

- Limitation: Discontinuities avoided but model falls outside the standard exponential class of models.
- MLE not feasible, but QML/NLS/GMM is feasible.

25. Linear feedback model

$$\begin{aligned}\lambda_{it} &= \rho y_{it-1} + \exp(\mathbf{x}'_{1it}\boldsymbol{\beta}_1 + + \mathbf{x}'_{\mathbf{2it}}\boldsymbol{\beta}_{\mathbf{2}} + \gamma_1 y_{i0} + \mathbf{z}'_i\gamma_2 + w_i) \\ &= \rho y_{it-1} + \exp(w_i)\exp(\mathbf{x}'_{1it}\boldsymbol{\beta}_1 + + \mathbf{x}'_{\mathbf{2it}}\boldsymbol{\beta}_{\mathbf{2}} + \gamma_1 y_{i0} + \mathbf{z}'_i\gamma_2 + w_i)\end{aligned}$$

- MLE not feasible because the functional form is no longer belongs in the exponential family.
- GMM which uses differencing transformations will eliminate initial values and correlated heterogeneity.
- NLS method for estimation can identify the conditional mean function under certain conditions.

$$\min_{\{\rho,\boldsymbol{\beta},\gamma\}} \frac{1}{2NT} \sum (y_{it} - \lambda_{it})^2$$

- To allow for a RE type extension should use a robust estimator of the covariance matrix.

Example: EFM vs LFM

```
. * Linear Feedback Model with Initial Condition Control
. gmm (officevis - {rho}*L.officevis - exp({xb: T0officevis insprv educ age income t
>  instruments(L.officevis T0officevis insprv educ age income totchr) onestep vce(cl

Final GMM criterion Q(b) =  7.35e-23

GMM estimation

Number of parameters =   8
Number of moments    =   8
Initial weight matrix: Unadjusted                     Number of obs   =   69027

                           (Std. Err. adjusted for 9861 clusters in dupersid)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        /rho |   .5366234   .0248079    21.63   0.000     .4880008     .585246
/xb_T0offi~s |   .0672159   .0038061    17.66   0.000     .0597561    .0746758
  /xb_insprv |   .1509578   .0408185     3.70   0.000     .0709551    .2309606
    /xb_educ |   .0375916   .0062318     6.03   0.000     .0253774    .0498058
     /xb_age |   .1234875   .0119579    10.33   0.000     .1000504    .1469245
  /xb_income |  -.0002804   .0006164    -0.45   0.649    -.0014885    .0009277
  /xb_totchr |   .3270725   .0154421    21.18   0.000     .2968066    .3573383
         /b0 |  -2.187085    .096687   -22.62   0.000    -2.376588   -1.997582
------------------------------------------------------------------------------
Instruments for equation 1: L.officevis T0officevis insprv educ age income totchr _c
```

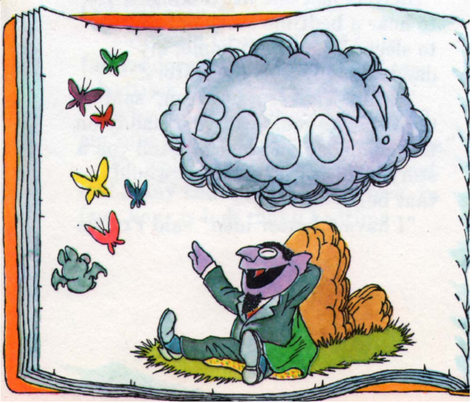
More on LFM vs. EFM

- Sensitivity to omitted $\mathbf{y}_0$ and $\mathbf{z}$ varies between LFM and EFM
- Monte Carlo analysis suggests omission leads to biases especially in the coefficient of lagged variable.
- EFM is preferred on predictive performance when the proportion of zeros is high.
- LFM does better when the mean of y is high and proportion of zeros small.
- NLS turns out to be a robust estimator for the LFM. Should be considered as a serious alternative for count panel models under certain conditions.

Concluding Remarks

- Much progress in estimating panel count models, especially in dealing with endogeneity and nonseprable heterogeneity.
- Great progress in variance estimation.
- RE models pose fewer problems.
- For FE models moment-based/IV methods seem more tractable for handling endogeneity and dynamics. Stata's new suite of GMM commands are very helpful in this regard.
- Because FE models do not currently handle important cases, and have other limitations, CCR panel model with initial conditions, is an attractive alternative, at least for balanced panels.

For Count Data magic, if you don't have the Thundercloud,



use STATA® 11 instead!

References

Hausman, J.A., B.H. Hall and Z. Griliches (1984), "Econometric Models for Count Data With an Application to the Patents-R and D Relationship," *Econometrica*, 52, 909-938.

Chamberlain, G. (1984). Panel Data. In Handbook of Econometrics, Volume II, ed. by Z. Griliches and M. Intriligator, 1247-1318. Amsterdam: North-Holland.

Wooldridge, J. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20, 39-54.

Chernozhukov, V; Fernandez-Val, I.; Kowalski, AE (2009) Censored Quantile Instrumental Variable Estimation via Control Functions. Discussion paper.

Koenker, R. (2004) Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74 – 89

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46, 69-85.

Stata Release 11 Manuals

Windmeijer, F.A.G. (2008), "GMM for Panel Count Data Models," ch.18 in L. Matyas and P. Sivestre eds., *The Econometrics of Panel Data*, Springer.