Advances in Count Data Regression: III. Time series and panel data

A. Colin Cameron
Univ. of Calif. - Davis

28th Annual Workshop in Applied Statistics
Southern California Chapter of the American Statistical Association
Held at University of California - Los Angeles
http://cameron.econ.ucdavis.edu/racd/count.html

March 28 2009

Outline of all Lectures

I. Basic cross-section methods:
- Poisson, GLM, negative binomial

II. More advanced cross-section methods:
- Hurdle, zero-inflated, finite mixtures, endogeneity

III. Time series and panel methods

IV. Further Topics:
- multivariate, maximum simulated likelihood

Outline of beyond cross-section count

- Introduction
- Time series data
- Panel data

Time series data

- Data $(y_t, \mathbf{x}_t)$ with
  Histories $\mathbf{y}^{(t-1)} = (y_{t-1}, y_{t-2}, ...)$ and $\mathbf{X}^{(t)} = (\mathbf{x}_t, \mathbf{x}_{t-1}, ...)$
- Many different models exist and there is no clear preferred model.
- Distinction between
  - Observation-driven models: time series dependence by direct dependence of moments or density on past outcomes
    e.g. $y_t = \rho y_{t-1} + \mathbf{x}_t'\beta + \varepsilon_t$
  - Parameter-driven models: time series dependence induced by latent variable process
    e.g. $y_t = \mathbf{x}_t'\beta + u_t$ and $u_t = \rho u_{t-1} + \varepsilon_t$
- Recent surveys:
  - Jung, Kukuk and Liesenfeld (2006)
  - Davis, Dunsmuir, Streett (2003)

Integer-valued ARMA (INARMA)

- Observation-driven approach that is most appealing theoretically
  - extends linear AR(1): $y_t = \rho y_{t-1} + \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t$
  - but difficult to implement
- INAR(1) process (no regressors) for integer counts is

$$y_t = \rho \circ y_{t-1} + \varepsilon_t, \qquad 0 \leq \rho < 1,$$

  - $\circ$ is the binomial thinning operator with $\rho \circ y = \sum_{j=1}^{y} u_j$ where $\Pr[u_j = 1] = p$ and $\Pr[u_j = 1] = 1 - p$
  - ε_t is an i.i.d. latent count variable, e.g. Poisson
  - Al-Osh and Alzaid (1987), McKenzie (1986).
- Properties include

$$\mathrm{E}[y_t|y_{t-1}] = \rho y_{t-1} + \mathrm{E}[\varepsilon_t] .$$

- Regression case

$$\begin{aligned} y_t &= \rho_t \circ y_{t-1} + \varepsilon_t \\ \rho_t &= \exp(\mathbf{z}_t'\gamma)/(1+\exp(\mathbf{z}_t'\gamma)) \\ \varepsilon_t &\sim \mathit{Poisson}[\exp(\mathbf{x}_t'\boldsymbol{\beta})] \end{aligned}$$

- Estimate by NLS or GMM using moment conditions is fine (Brannas (1995)).
- Estimate by MLE is difficult.
- Can extend to integer ARMA and ε_t negative binomial.

Observation-driven approach

- Observation-driven approach that incorporates lags into μ_t.
- Simplest approaches for AR(1) fail.
- The following model is explosive for $\rho > 0$

$$\begin{aligned} y_t &\sim \mathcal{P}[\exp(\mathbf{x}_t'\boldsymbol{\beta} + \rho y_{t-1})] \\ &\sim \mathcal{P}[\exp(\mathbf{x}_t'\boldsymbol{\beta}) \times e^{\rho y_{t-1}}] \end{aligned}$$

- The following model has problem if $y_{t-1} = 0$

$$y_t \sim \mathcal{P}[\exp(\mathbf{x}_t'\boldsymbol{\beta} + \rho \ln y_{t-1})]$$

  - Zeger and Qaqish (1988) proposed use $y_{t-1}^* = \min(c, y_{t-1})$.
- Leads to alternative models, including the following.

- Autoregressive conditional Poisson (Heinen (2003))
  - Simple example:

$$\begin{aligned} y_t|\mathbf{y}^{(t-1)} &\sim \mathit{Poisson}[\mu_t] \\ \mu_t &= \mathrm{E}[y_t|\mathbf{y}^{(t-1)}] = \omega + \alpha_1 \mu_{t-1} \end{aligned}$$

  - Regressors included by replace μ_t by $\mu_t^* = \mu_t \exp(\mathbf{x}_t'\boldsymbol{\beta})$
  - More generally can have ARMA versions and GARCH.
- Generalized linear autoregressive moving average model (Davis (1999)
  - Simple example

$$\begin{aligned} y_t|\mathbf{y}^{(t-1)} &\sim \mathit{Poisson}[\mu_t] \\ \ln \mu_t &= \omega + \alpha_1 (y_{t-1} - \mu_{t-1})/\mu_{t-1}^{\rho} \end{aligned}$$

  - Regressors included by replace μ_t by $\mu_t^* = \mu_t \exp(\mathbf{x}_t'\boldsymbol{\beta})$
  - More generally can have ARMA versions.

Parameter-driven approach

- Include a multiplicative serially-correlated latent variable to the Poisson mean.
- For example Poisson with AR(1) latent variable

$$\begin{aligned} y_t &= Poisson[\exp(\mathbf{x}_t'\boldsymbol{\beta}) \times u_t] \\ \ln u_t &= \rho \ln u_{t-1} + \varepsilon_t, \quad |\rho| < 1 \\ \varepsilon_t &\sim \mathcal{N}[0,1] \end{aligned}$$

- Estimate by quasi-likelihood GEE-type approach (Zeger (1988))
  - F.o.c. are $\mathbf{D}'\mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu}(\boldsymbol{\beta})) = \mathbf{0}$ where $\mathbf{V}^{-1} \simeq \mathrm{Var}[\mathbf{y}]^{-1}$.
- Estimate by MLE is difficult (high-dimensional integral)
  - For example, use Markov Chain with efficient importance sampling (Jung et al. (2006)).

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 $N \to \infty$ and T small
  - observations are uncorrelated across individuals (i)
  - observations may be correlated over time (t) for given individual
  - for most estimators panel can be unbalanced.
- Economics focuses on case that $\mathrm{Cov}[\alpha_i, \mathbf{x}_{it}] \neq 0$
  - then pooled OLS is inconsistent.

Linear panel: pooled OLS and pooled FGLS

- Assume α_i is independent of $\mathbf{x}_{it}$ with mean 0 (so part of error)

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

- Pooled OLS of y_{it} on x_{it} gives consistent $\boldsymbol{\beta}$.
  - Get cluster-robust standard errors where cluster on the individual.
- Pooled FGLS
  - Assume a model for $\mathrm{Cor}[\alpha_i + \varepsilon_{it}, \alpha_i + \varepsilon_{is}]$ e.g. equicorrelation.
  - Do FGLS of y_{it} on x_{it} to improve efficiency
  - Can then get cluster-robust standard errors

$$\widehat{\mathrm{V}}_{\mathrm{ROB}}[\widehat{\beta}] = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_i \sum_t (y_{it} - \mathbf{x}_{it}'\widehat{\beta})^2 \mathbf{x}_{it}\mathbf{x}_{it}'\right)(\mathbf{X}'\mathbf{X})^{-1}$$

.

- Also called population-averaged:
  - essential assumption is $\mathrm{E}[y_{it}|\mathbf{x}_{it}] = \mathbf{x}_{it}'\boldsymbol{\beta}$.

Linear panel: random effects estimator

- Again: $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + (\alpha_i + \varepsilon_{it})$.
- Assume α_i is i.i.d. $[0, \sigma_\alpha^2]$ and ε_{it} is i.i.d. $[0, \sigma_\varepsilon^2]$.
- Then GLS is OLS in the transformed model (with $\theta_i = 1 - \sqrt{\sigma_\varepsilon^2/[T_i\sigma_\alpha^2 + \sigma_\varepsilon^2]}$)

$$(y_{it} - \theta_i\bar{y}_i) = (\mathbf{x}_{it} - \theta_i\bar{\mathbf{x}}_i)'\boldsymbol{\beta} + \text{ i.i.d. error}.$$

- FGLS is OLS of $(y_{it} - \widehat{\theta}_i\bar{y}_i)$ on $(\mathbf{x}_{it} - \widehat{\theta}_i\bar{\mathbf{x}}_i)'\boldsymbol{\beta}$.
  - Can get cluster-robust standard errors in case α_i and ε_{it} not i.i.d.
  - essential assumption is $\mathrm{E}[y_{it}|\mathbf{x}_{it}] = \mathbf{x}_{it}'\boldsymbol{\beta}$.
- Mixed model or hierarchical linear model also allows slopes $\boldsymbol{\beta}_i$ to be random.

Linear panel: within or fixed effects estimator

- Again: $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + (\alpha_i + \varepsilon_{it})$.
- Assume α_i is potentially correlated with $\mathbf{x}_{it}$
  - e.g. Earnings regression and α_i is time-invariant unobserved ability.
- Eliminate α_i by mean-differencing

$$(y_{it} - \bar{y}_i) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (u_{it} - \bar{u}_i).$$

- Within estimator is OLS of $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ (with no intercept).
  - Can get cluster-robust standard errors.

Linear panel: first differences estimator

- Again: $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + (\alpha_i + \varepsilon_{it})$.
- Eliminate α_i by first-differencing

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

- First difference estimator is OLS of $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ (with no intercept).
  - Can get cluster-robust standard errors.
  - Can adapt this to IV for dynamic models with lagged dependent variable as regressors (Arellano-Bond).

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: $\widehat{\mathrm{V}}[y] = 4.50^2 \simeq 7 \times \bar{y}$.

```
. 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
```

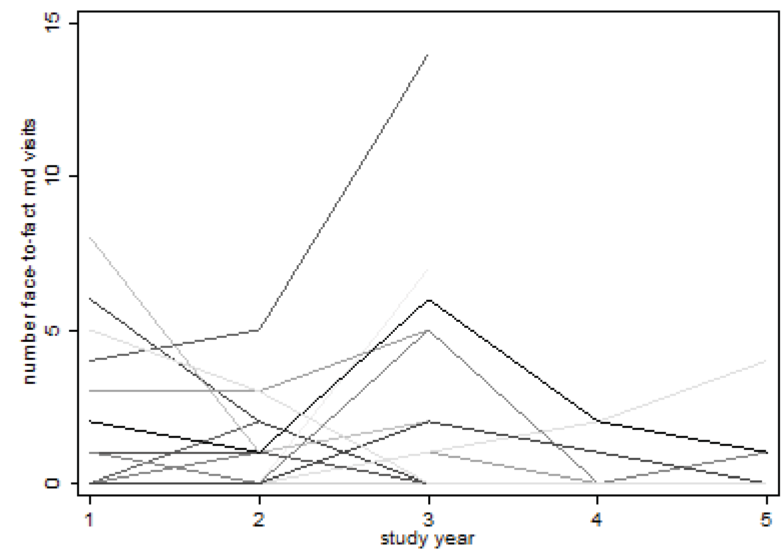
Panel is unbalanced. Most are in for 3 years or 5 years.

```
. 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
```


Time series plots for the first 20 individuals. Much serial correlation.



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- Poisson random effects model is

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

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)
  - $\text{E}[y_{it}|\mathbf{x}_{it}, \beta] = \exp(\mathbf{x}_{it}'\beta)$
- 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)
  - $\text{E}[y_{it}|\mathbf{x}_{it}, \beta] = \exp(\mathbf{x}_{it}'\beta)$ 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
```

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 $N \to \infty$.
  - but no problem as can eliminate α_i.
- Eliminate α_i by quasi-differencing as follows

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

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

- The final result implies

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

- Poisson fixed effects estimator solves the corresponding sample moment conditions

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

  - Get cluster-robust standard errors.
- Consistency requires

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

- This 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.

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

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

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 different Poisson panel 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					
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.

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: One possibility 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: fixed effects

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

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

- And use an alternative quasi-difference

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

- 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.

References

- Time Series:
  - Jung, R.C., M. Kukuk, and R. Liesenfeld (2006), "Time Series of Count Data: Modeling, Estimation and Diagnostics," *Computational Statistics and Data Analysis*, 51, 2350-2364.
  - Davis, Dunsmuir, Streett (2003), "Observation-driven Models for Time Series Models involving Counts," *Biometrika*, 90, 777-790.
- Panel Data:
  - 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.
  - 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.