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 \((y_t, x_t)\) with
- Histories \(y^{(t-1)} = (y_{t-1}, y_{t-2}, \ldots)\) and \(X^{(t)} = (x_t, x_{t-1}, \ldots)\)
- 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. \(y_t = \rho y_{t-1} + x_t^\beta + \varepsilon_t\)
  - Parameter-driven models: time series dependence induced by latent
    variable process
    e.g. \(y_t = x_t^\beta + u_t\) 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): \( y_t = \rho y_{t-1} + x'_t \beta + \varepsilon_t \)
  - but difficult to implement
- INAR(1) process (no regressors) for integer counts is
  \[ y_t = \rho \circ y_{t-1} + \varepsilon_t, \quad 0 \leq \rho < 1, \]
  - \( \circ \) is the binomial thinning operator with
    \( \rho \circ y = \sum_{j=1}^{y} u_j \) where \( \Pr[u_j = 1] = \rho \) and \( \Pr[u_j = 1] = 1 - \rho \)
  - \( \varepsilon_t \) is an i.i.d. latent count variable, e.g. Poisson
- Properties include
  \[ E[y_t|y_{t-1}] = \rho y_{t-1} + E[\varepsilon_t]. \]

Observation-driven approach

- Observation-driven approach that incorporates lags into \( \mu_t \).
- Simplest approaches for AR(1) fail.
- The following model is explosive for \( \rho > 0 \)
  \[ y_t \sim \mathcal{P}\left[\exp(x'_t \beta + \rho y_{t-1})\right] \sim \mathcal{P}\left[\exp(x'_t \beta) \times e^{\rho y_{t-1}}\right] \]
- The following model has problem if \( y_{t-1} = 0 \)
  \[ y_t \sim \mathcal{P}\left[\exp(x'_t \beta + \rho \ln y_{t-1})\right] \]
  - Zeger and Qaqish (1988) proposed use \( y'_{t-1} = \min(c, y_{t-1}) \).
- Leads to alternative models, including the following.

Regression case

\[ y_t = \rho_t \circ y_{t-1} + \varepsilon_t \]
\[ \rho_t = \exp(z'_t \gamma)/(1 + \exp(z'_t \gamma)) \]
\[ \varepsilon_t \sim \text{Poisson}[\exp(x'_t \beta)] \]
- Estimate by NLS or GMM using moment conditions is fine (Brannas (1995)).
- Estimate by MLE is difficult.
- Can extend to integer ARMA and \( \varepsilon_t \) negative binomial.

Autoregressive conditional Poisson (Heinen (2003))

- Simple example:
  \[ y_t|y^{(t-1)} \sim \text{Poisson}[\mu_t] \]
  \[ \mu_t = E[y_t|y^{(t-1)}] = \omega + \alpha_1 \mu_{t-1} \]
- Regressors included by replace \( \mu_t \) by \( \mu^*_t = \mu_t \exp(x'_t \beta) \)
- More generally can have ARMA versions and GARCH.
- Generalized linear autoregressive moving average model (Davis (1999))
  - Simple example
    \[ y_t|y^{(t-1)} \sim \text{Poisson}[\mu_t] \]
    \[ \ln \mu_t = \omega + \alpha_1 (y_{t-1} - \mu_{t-1})/\mu^\rho_{t-1} \]
  - Regressors included by replace \( \mu_t \) by \( \mu^*_t = \mu_t \exp(x'_t \beta) \)
  - 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

\[ y_t = \text{Poisson}[\exp(x_t' \beta) \times u_t] \]
\[ \ln u_t = \rho \ln u_{t-1} + \varepsilon_t, \quad |\rho| < 1 \]
\[ \varepsilon_t \sim \mathcal{N}[0,1] \]

- Estimate by quasi-likelihood GEE-type approach (Zeger (1988))
  - F.o.c. are \( D' \tilde{V}^{-1} (y - \mu(\beta)) = 0 \) where \( \tilde{V}^{-1} \approx \text{Var}[y]^{-1} \).
- 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

\[ y_{it} = x_{it}' \beta + \alpha_i + \varepsilon_{it}, \quad i = 1, ..., N, \ t = 1, ..., T. \]
  - So different people have different unobserved intercepts \( \alpha_i \).
- Goal is to consistently estimate slope parameters \( \beta \).
- 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 \( \text{Cov}[\alpha_i, x_{it}] \neq 0 \)
  - then pooled OLS is inconsistent.

Linear panel: pooled OLS and pooled FGLS

- Assume \( \alpha_i \) is independent of \( x_{it} \) with mean 0 (so part of error)

\[ y_{it} = x_{it}' \beta + (\alpha_i + \varepsilon_{it}). \]

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

\[ \hat{V}_{\text{ROB}}[\hat{\beta}] = (XX)^{-1} \left( \sum_t \sum_i (y_{it} - x_{it}' \hat{\beta})^2 x_{it} x_{it}' \right) (XX)^{-1} \]

- Also called population-averaged:
  - essential assumption is \( \text{E}[y_{it} | x_{it}] = x_{it}' \beta \).

Linear panel: random effects estimator

- Again: \( y_{it} = x_{it}' \beta + (\alpha_i + \varepsilon_{it}) \).
- Assume \( \alpha_i \) is i.i.d. \( [0, \sigma^2_{\alpha}] \) and \( \varepsilon_{it} \) is i.i.d. \( [0, \sigma^2] \).
- Then GLS is OLS in the transformed model (with \( \theta_i = 1 - \sqrt{\sigma^2 / (T \sigma^2_{\alpha} + \sigma^2)} \))

\[ (y_{it} - \theta_i \bar{y}_i) = (x_{it} - \theta_i \bar{x}_i)' \beta + \text{i.i.d. error}. \]

- FGLS is OLS of \( (y_{it} - \theta_i \bar{y}_i) \) on \( (x_{it} - \theta_i \bar{x}_i)' \beta \).
  - Can get cluster-robust standard errors in case \( \alpha_i \) and \( \varepsilon_{it} \) not i.i.d.
  - essential assumption is \( \text{E}[y_{it} | x_{it}] = x_{it}' \beta \).
- Mixed model or hierarchical linear model also allows slopes \( \beta_i \) to be random.
Linear panel: within or fixed effects estimator

- Again: $y_{it} = \mathbf{x}'_{it}\beta + (\alpha_i + \epsilon_{it})$.
- Assume $\alpha_i$ is potentially correlated with $\mathbf{x}_{it}$
  - e.g. Earnings regression and $\alpha_i$ is time-invariant unobserved ability.
- Eliminate $\alpha_i$ by mean-differencing
  $$(y_{it} - \bar{y}_i) = (\mathbf{x}_t - \bar{\mathbf{x}}_i)'\beta + (u_t - \bar{u}_i).$$
- Within estimator is OLS of $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_t - \bar{\mathbf{x}}_i)$ (with no intercept).
  - Can get cluster-robust standard errors.

Panel counts: data example

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

```
use mus18data.dta, clear
describe mdu l1coins ndisease female age l1fam child id year
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mdu</td>
<td>20185</td>
<td>2.850696</td>
<td>4.504765</td>
<td>0</td>
<td>77</td>
</tr>
<tr>
<td>l1coins</td>
<td>20185</td>
<td>2.383588</td>
<td>2.041713</td>
<td>0</td>
<td>4.563488</td>
</tr>
<tr>
<td>ndisease</td>
<td>20185</td>
<td>11.2445</td>
<td>6.741674</td>
<td>0</td>
<td>58.6</td>
</tr>
<tr>
<td>female</td>
<td>20185</td>
<td>.5159424</td>
<td>4.97252</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>age</td>
<td>20185</td>
<td>25.2114</td>
<td>16.76759</td>
<td>0</td>
<td>64.27515</td>
</tr>
</tbody>
</table>

| l1fam    | 20185| 1.248404 | .5390681  | 0    | 2.639057  |
| child    | 20185| .4804168 | .4903972  | 0    | 1      |
| id       | 20185| 357971.2 | 1.808856  | 125024 | 632167  |
| year     | 20185| 2.420044 | 1.217237  | 1    | 5      |

Dependent variable mdu is very overdispersed: $\hat{\text{V}}[y] = 4.50^2 \simeq 7 \times \bar{y}$.
Panel is unbalanced. Most are in for 3 years or 5 years.

```
xtdescribe
   id:  125024, 125025, ..., 632167  n =  906
   year: 1, 2, ..., 5  t =  5
   Delta(year) = 1 unit
   Span(year) = 5 periods
   (10th 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.

```
xpanel summary of dependent variable
xtsuen mdu
Variable | Mean  Std. Dev.  Min  Max  Observations
----------|------|------|------|------|-------------------
        mdu overall | 2.86096  4.504765  0  77  N = 20186
         between | 3.785971  63.33333  0  63.33333  n = 5908
         within | 2.575881  -34.47264  40.0667  T-bar = 3.42672
```

Only time-varying regressors are age, lfin 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 \( y_{it}|x_{it}, \beta \sim \text{Poisson}[\exp(x'_{it}\beta)] \)
- Pooled Poisson of \( y_{it} \) on intercept and \( x_{it} \) gives consistent \( \beta \).
  - But get cluster-robust standard errors where cluster on the individual.
  - These control for both overdispersion and correlation over \( t \) for given \( i \).

Panel Poisson method 2: population-averaged

- Assume that for the \( i^{th} \) observation moments are like for GLM Poisson:
  \[
  E[y_{it}|x_{it}] = \exp(x'_{it}\beta) \\
  V[y_{it}|x_{it}] = \phi \times \exp(x'_{it}\beta).
  \]

- Stack the conditional means for the \( i^{th} \) individual:
  \[
  E[y_{i}|X_{i}] = m_{i}(\beta) = \begin{bmatrix} \exp(x'_{i1}\beta) \\ \vdots \\ \exp(x'_{iT}\beta) \end{bmatrix}.
  \]
  where \( y_{i} = [y_{i1}, \ldots, y_{iT}]' \) and \( X_{i} = [x_{i1}, \ldots, x_{iT}]' \).

- Stack the conditional variances for the \( i^{th} \) individual.
  - With no correlation
    \[
    V[y_{i}|X_{i}] = \phi H_{i}(\beta) = \phi \times \text{Diag}[\exp(x'_{it}\beta)].
    \]

Pooled Poisson with cluster-robust standard errors

- Pooled Poisson estimator with cluster-robust standard errors:
  \[
  \text{poisson mdv lcoins ndisease female age 1fam child, vc(cluster id)}
  \]

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

  Poisson regression

| mdu    | Coef. | Robust Std. Err. | z     | P>|z| | [95% Conf. Interval]        |
|--------|-------|------------------|-------|-----|-----------------------------|
| lcoins | -0.080423 | 0.080693 | -10.10 | 0.000000 | -0.191587 | -0.180241 |
| ndisease | 0.118914 | 0.0 36024 | 11.04 | 0.000000 | 0.100371 | -0.059804 |
| female | 0.177852 | 0.0 34251 | 5.01 | 0.000000 | 0.170637 | -0.020282 |
| age    | -0.040585 | 0.018291 | 2.40 | 0.016667 | -0.067860 | 0.007715 |
| 1fam   | -1.419981 | 0.234314 | -4.58 | 0.000000 | -2.877534 | -0.962426 |
| child  | 0.106045 | 0.050690 | 2.03 | 0.042250 | -0.013844 | -0.233542 |
| _cons  | 1.478789 | 0.078573 | 9.53 | 0.000000 | 1.320758 | 1.636822 |

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!!
The GLM estimator solves: \[ \sum_{i=1}^{N} \frac{\partial m_i'(\beta)}{\partial \beta} H_i(\beta)^{-1} (y_i - m_i(\beta)) = 0. \]

Generalized estimating equations (GEE) estimator or population-averaged estimator (PA) of Liang and Zeger (1986) solves

\[ \sum_{i=1}^{N} \frac{\partial m_i'(\beta)}{\partial \beta} \hat{\Omega}_i^{-1} (y_i - m_i(\beta)) = 0, \]

where \( \hat{\Omega}_i \) equals \( \Omega_i \) in with \( R(\alpha) \) replaced by \( R(\hat{\alpha}) \) where \( \text{plim} \hat{\alpha} = \alpha \).

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

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

where \( \hat{D}_g = \frac{\partial m_g'(\beta)}{\partial \beta} |_{\hat{\beta}}, \hat{D} = [\hat{D}_1, \ldots, \hat{D}_G]', \hat{u}_g = y_g - m_g(\hat{\beta}), \)

and now \( \hat{\Omega}_g = H_g(\hat{\beta})^{-1/2} R(\hat{\rho}) H_g(\hat{\beta})^{1/2} \).

- The asymptotic theory requires that \( G \to \infty \).

### Population-averaged Poisson with unstructured correlation

<table>
<thead>
<tr>
<th>GEE population-averaged model</th>
<th>Number of obs</th>
<th>20,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group and time vars:</td>
<td>id: year</td>
<td>3098</td>
</tr>
<tr>
<td>Link: log</td>
<td>Obs per group</td>
<td>1</td>
</tr>
<tr>
<td>Family: Poisson</td>
<td>avg</td>
<td>3.4</td>
</tr>
<tr>
<td>Correlation: unstructured</td>
<td>max</td>
<td>508.61</td>
</tr>
<tr>
<td>Scale parameter:</td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|  | Coef. | Semi-robust | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|  |       |             |           |      |     |                    |
| lnc | -0.09444 | 0.007782 | -10.24 | 0.000 | <.0596 | -0.0652 | 0.0094 |
| ncdise | 0.0346 | 0.00242 | 14.28 | 0.000 | <.0288561 | 0.0327 | 0.0366 |
| female | 0.1585 | 0.0334 | 4.74 | 0.000 | <.0929649 | 0.2424 | 0.0756 |
|  | 0.0099 | 0.0015 | 2.01 | 0.000 | <.0001083 | 0.0001 | 0.0002 |
|  | -0.1465 | 0.0293 | -4.99 | 0.000 | <.0182135 | 0.0360 | 0.0166 |
|  | 0.1013677 | 0.04301 | 2.36 | 0.008 | <.0170966 | 0.0343 | 0.1758 |
|  | -0.7764 | 0.0712 | -10.83 | 0.000 | <.0001083 | 0.065897 | 0.917054 |

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.

### Panel Poisson method 3: random effects

- Poisson random effects model is

\[
y_{it} | x_{it}, \beta, \alpha_i \sim \text{Pois}[\exp(x_{it}'\beta)] \sim \text{Pois}[\exp(\ln \alpha_i + x_{it}'\beta)]
\]

where \( \alpha_i \) is unobserved but is not correlated with \( x_{it} \).

- RE estimator 1: Assume \( \alpha_i \) is \( \text{Gamma}[1, \eta] \) distributed
  - closed-form solution exists (negative binomial)
  - \( E[y_{it}|x_{it}, \beta] = \exp(x_{it}'\beta) \)

- RE estimator 2: Assume \( \ln \alpha_i \) is \( \mathcal{N}[0, \sigma^2] \) distributed
  - closed-form solution does not exist (one-dimensional integral)
  - can extend to slope coefficients (higher-dimensional integral)
  - \( E[y_{it}|x_{it}, \beta] = \exp(x_{it}'\beta) \) aside from translation of intercept.
Poisson random effects (gamma) with panel bootstrap se’s

Random-effects Poisson regression
Number of obs: 20186
Number of groups: 5908

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

Log likelihood: -43240.556
Wald chi2(6) = 529.10
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 |
|-----|----------------|---------------------|---|------|-------------------------------|
| Licon | -0.0687258 | 0.0060972 | -10.20 | 0.000 | -0.1047004, -0.078511 |
| female | 0.191807 | 0.0269294 | 7.14 | 0.000 | 0.138929, 0.244886 |
| age | 0.0915159 | 0.0065242 | 1.41 | 0.000 | 0.078675, 0.104955 |
| child | 0.108927 | 0.0495487 | 2.19 | 0.029 | 0.011151, 0.205186 |

/lnalpha | 0.0251258 | 0.0202518 | -0.12 | 0.898 | 0.009754, 0.05909 |

alpha | 1.025444 | 0.027175 | 3.82 | 0.000 | 0.918326, 1.13256 |

Likelihood-ratio test of alpha=0: chi2(1) = 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
  \[ y_{it} | x_{it}, \beta, \alpha_i \sim \text{Pois} [\alpha_i \exp(x'_{it}\beta)] \sim \text{Pois} [\exp(\ln \alpha_i + x'_{it}\beta)] \]
  where \( \alpha_i \) is unobserved and is possibly correlated with \( x_{it} \).
- In theory need to estimate \( \beta \) and \( \alpha_1, \ldots, \alpha_N \).
  - potential incidental parameters problem \( N + K \) parameters and \( NT \) observations with \( N \to \infty \).
  - but no problem as can eliminate \( \alpha_i \).
- Eliminate \( \alpha_i \) by quasi-differencing as follows
  \[
  \begin{align*}
  E[y_{it} | x_{i1}, \ldots, x_{iT}, \alpha_i] & = \alpha_i \lambda_t \\
  \lambda_t & = \exp(x'_{it}\beta) \\
  \Rightarrow & \ E[y_{it} | x_{i1}, \ldots, x_{iT}, \alpha_i] = \alpha_i \lambda_t \\
  \Rightarrow & \ E \left[ \left( y_{it} - \frac{\lambda_t}{\bar{\lambda}_i} \bar{y}_i \right) | x_{i1}, \ldots, x_{iT} \right] = 0
  \end{align*}
  \]
  - The first line assumes regressors \( x_{it} \) are strictly exogenous.
  - This is stronger than weakly exogenous.

- This estimator for \( \beta \) can also be obtained in the following ways under
  fully parametric assumption that
  \[ y_{it} | x_{it}, \beta, \alpha_i \sim \text{Pois} [\alpha_i \exp(x'_{it}\beta)] \]
  - 1. Obtain the MLE of \( \beta \) and \( \alpha_1, \ldots, \alpha_N \).
  - 2. Obtain the conditional MLE based on the conditional density
    \[
    f(y_{i1}, \ldots, y_{iT} | \bar{y}_i, x_{i1}, \ldots, x_{iT}, \beta, \alpha_i) = \frac{\prod_{t=1}^T f(y_{it} | x_{it}, \beta, \alpha_i)}{f(\bar{y}_i | x_{i1}, \ldots, x_{iT}, \beta, \alpha_i)}
    \]
    - But should then use cluster-robust standard errors and not default ML se’s.
Poisson fixed effects with panel bootstrap s.e.'s

. xtpoisson mdumoons ndisease female age t1fam child, fe vce(boot, reps(100) seed(10)
(running xtpoisson on estimation sample)

Bootstrap replications (100)

<table>
<thead>
<tr>
<th>Replication</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
</table>

| Number of obs | 1791 |
| Number of groups | 4977 |

| Obs per group | min | 2 | avg | 3.6 | max | 5 |

Log likelihood = -24173.211

H0: chi2(2) = 1.64
Prob > chi2 = 0.2002

(Replications based on 4977 clusters in id)

| Variable | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
|-----------|-------|-----------|---|-----|-----------------|
| age       | -0.012009 | 0.009077 | -1.18 | 0.239 | -0.028169 | 0.004739 |
| t1fam     | 0.0877134 | 0.1125783 | 0.78 | 0.436 | -0.132936 | 0.3083627 |
| child     | 0.159867 | 0.0738452 | 1.44 | 0.151 | -0.038747 | 0.250706 |

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

Panel Poisson: estimator comparison

- Strength of fixed effects versus random effects
  - Allows $a_i$ to be correlated with $x_{it}$.
  - 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 $a_i$:

$$ME_j = \partial E[y_{it}/x_{it}, j = a_i \exp(x_{it}'B)\beta_j$$

and $a_i$ is unknown.

Comparison of different Poisson panel estimators with cluster-robust s.e.'s

<table>
<thead>
<tr>
<th>Variable</th>
<th>POOLED</th>
<th>POPAVE</th>
<th>RE_GAMMA</th>
<th>RE_NOR-L</th>
<th>FIXED</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1coins</td>
<td>-0.0809</td>
<td>-0.0804</td>
<td>-0.0797</td>
<td>-0.0115</td>
<td></td>
</tr>
<tr>
<td>ndisease</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0073</td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>0.0339</td>
<td>0.0346</td>
<td>0.0368</td>
<td>0.0469</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>0.1718</td>
<td>0.1585</td>
<td>0.1667</td>
<td>0.2084</td>
<td></td>
</tr>
<tr>
<td>t1fam</td>
<td>0.0041</td>
<td>0.0031</td>
<td>0.0019</td>
<td>0.0027</td>
<td></td>
</tr>
<tr>
<td>child</td>
<td>0.0087</td>
<td>0.0015</td>
<td>0.0016</td>
<td>0.0012</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>-0.1462</td>
<td>-0.1467</td>
<td>-0.1352</td>
<td>-0.1443</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>-0.0375</td>
<td>-0.0304</td>
<td>0.0069</td>
<td>0.0261</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>0.1039</td>
<td>0.1014</td>
<td>0.1033</td>
<td>0.0757</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>0.0507</td>
<td>0.0430</td>
<td>0.0092</td>
<td>0.0945</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>0.7488</td>
<td>0.7765</td>
<td>0.7574</td>
<td>0.2873</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>0.0708</td>
<td>0.0717</td>
<td>0.0755</td>
<td>0.0642</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>$11$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$20136$</td>
<td>-625.9</td>
</tr>
<tr>
<td>$20136$</td>
<td>-43241</td>
</tr>
<tr>
<td>$20136$</td>
<td>-43227</td>
</tr>
<tr>
<td>$20136$</td>
<td>-24173</td>
</tr>
</tbody>
</table>

Legend: b/s.e.
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 ($\alpha_i$)
  - e.g. High $\alpha_i$ 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:
  $$y_{it} = \alpha_i + \rho y_{i,t-1} + x_i' \beta + u_{it}.$$ 
- Poisson model: One possibility is
  $$\mu_{it} = \alpha_i \lambda_{it-1} = \alpha_i \exp(\rho y_{i,t-1}^{*} + x_i' \beta),$$
  $$y_{i,t-1}^{*} = \min(c, y_{i,t-1}).$$

Panel dynamic: fixed effects

- In fixed effects case Poisson FE estimator is now inconsistent.
- Instead assume weak exogeneity
  $$E [y_{it} | y_{i,t-1}, \ldots, y_{i1}, x_{i1}, \ldots, x_{it}] = \alpha_i \lambda_{it-1}.$$ 
- And use an alternative quasi-difference
  $$E \left[ (y_{it} - \lambda_{it-1} / \lambda_{it}) y_{i,t-1} | y_{i,t-1}, \ldots, y_{i1}, x_{i1}, \ldots, x_{it} \right] = 0.$$ 
- So MM or GMM based on
  $$E \left[ z_{it} \left( y_{it} - \frac{\lambda_{it-1}}{\lambda_{it}} y_{i,t-1} \right) \right] = 0$$
  where e.g. $z_{it} = (y_{i,t-1}, x_{i,t})$ in just-identified case.
- Windmeijer (2008) has recent discussion.

References

- Time Series:
- Panel Data: