Count Outcome Panel Data Models

230347 Advanced Microeconometrics
Tilburg University

Christoph Walsh

Count Data

- In this section, our outcome variable y_{it} only takes non-negative integer values: $y_{it} \in \{0, 1, 2, \dots\}$
- y_{it} is the number of times an event occurs for i during time period t.
- Examples:
  - Number of patents filed (measure of innovation).
  - Number of doctor visits.
  - Number of absent days (at work/school).
  - Number of accidents.

This Lecture

- In this lecture we will study:
  - The Poisson distribution.
  - Static fixed effects Poisson
  - Static random effects Poisson
  - Excess zeros in y_{it}.
  - Dynamic fixed effects Poisson

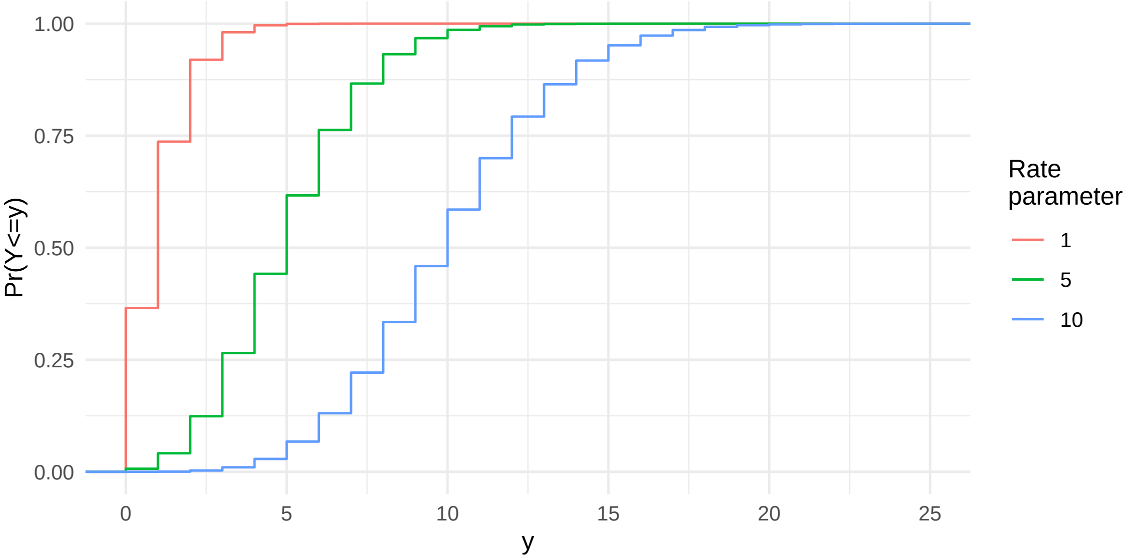
Poisson Distribution

- The Poisson distribution is often used to model count data.
- The discrete random variable Y counts the number of times an event occurs in one time period.
- If Y follows a Poisson distribution with rate $\mu > 0$, the probability of an event happening y times in a time period is:

$$\Pr(Y = y) = \frac{\exp(-\mu)\,\mu^y}{y!}$$

  where $y \in \{0, 1, 2, \dots\}$.
- The mean and variance of Y is μ.

CDF of the Poisson Distribution for $\mu \in \{1, 5, 10\}$



Static Fixed Effects Poisson

- We model y_{it} to be Poisson distributed with rate μ_{it}, where:

$$\mu_{it} = \alpha_i \exp\left(\mathbf{x}_{it}'\boldsymbol{\beta}\right) = \alpha_i \lambda_{it}$$

- $\alpha_i \geq 0$ is a multiplicative fixed effect.
  - Defining $\alpha_i = e^{\delta_i}$ gives $\mu_{it} = e^{\delta_i} \exp\left(\mathbf{x}_{it}'\boldsymbol{\beta}\right) = \exp\left(\delta_i + \mathbf{x}_{it}'\boldsymbol{\beta}\right)$
- In this model we can remove the fixed effect in three different ways:
  - Concentrating out α_i from the log likelihood.
  - A mean-differencing transformation.
  - Conditioning on a sufficient statistic ($\sum_{t=1}^{T} y_{it}$ here like in the logit case).
    - This third case is left as an exercise.

Likelihood

- Each observation's contribution to the likelihood is:

$$\Pr\left(y_{it}|\alpha_i, \boldsymbol{\beta}\right) = \frac{\exp\left(-\alpha_i \lambda_{it}\right)\left(\alpha_i \lambda_{it}\right)^{y_{it}}}{y_{it}!}$$

- If the y_{it} are iid, the i's contribution to the likelihood is:

$$\begin{aligned}
\Pr\left(y_{i1}, \ldots, y_{iT}|\alpha_i, \boldsymbol{\beta}\right) &= \prod_{t=1}^{T} \frac{\exp\left(-\alpha_i \lambda_{it}\right)\left(\alpha_i \lambda_{it}\right)^{y_{it}}}{y_{it}!} \\
&= \frac{\exp\left(-\alpha_i \sum_{t=1}^{T} \lambda_{it}\right) \prod_{t=1}^{T} \alpha_i^{y_{it}} \prod_{t=1}^{T} \lambda_{it}^{y_{it}}}{\prod_{t=1}^{T} y_{it}!}
\end{aligned}$$

Log Likelihood

The log likelihood for individual i is:

$$\log\left(\Pr\left(\boldsymbol{y}_i|\alpha_i,\boldsymbol{\beta}\right)\right) = \log\left(\frac{\exp\left(-\alpha_i\sum_{t=1}^{T}\lambda_{it}\right)\prod_{t=1}^{T}\alpha_i^{y_{it}}\prod_{t=1}^{T}\lambda_{it}^{y_{it}}}{\prod_{t=1}^{T}y_{it}!}\right)$$

$$= -\alpha_i\sum_{t=1}^{T}\lambda_{it} + \sum_{t=1}^{T}y_{it}\log\left(\alpha_i\right) + \sum_{t=1}^{T}y_{it}\log\left(\lambda_{it}\right) - \sum_{t=1}^{T}\log\left(y_{it}!\right)$$

- Taking the first-order condition with respect to α_i:

$$-\sum_{t=1}^{T}\lambda_{it} + \frac{\sum_{t=1}^{T}y_{it}}{\alpha_i} = 0 \qquad \Longrightarrow \qquad \alpha_i = \frac{\sum_{t=1}^{T}y_{it}}{\sum_{t=1}^{T}\lambda_{it}} \quad \forall i$$

Concentrated Log Likelihood

- Substituting $\alpha_i = \frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} \lambda_{it}}$ $\forall i$ into the log likelihood function:

$$\begin{aligned}
&\log\left(\mathcal{L}_{conc}\left(\boldsymbol{\beta}\right)\right) \\
&= \sum_{i=1}^{N}\left[-\alpha_i \sum_{t=1}^{T}\lambda_{it} + \sum_{t=1}^{T} y_{it}\log\left(\alpha_i\right) + \sum_{t=1}^{T} y_{it}\log\left(\lambda_{it}\right) - \sum_{t=1}^{T}\log\left(y_{it}!\right)\right] \\
&= \sum_{i=1}^{N}\left[-\sum_{t=1}^{T} y_{it} + \sum_{t=1}^{T} y_{it}\log\left(\frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T}\lambda_{it}}\right) + \sum_{t=1}^{T} y_{it}\log\left(\lambda_{it}\right) - \sum_{t=1}^{T}\log\left(y_{it}!\right)\right]
\end{aligned}$$

- Dropping constant terms leaves us with:

$$\begin{aligned}
\log\left(\mathcal{L}_{conc}\left(\boldsymbol{\beta}\right)\right) &\propto \sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}\log\left(\frac{\lambda_{it}}{\sum_{s=1}^{T}\lambda_{is}}\right) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}\log\left(\frac{\exp\left(\boldsymbol{x}_{it}'\boldsymbol{\beta}\right)}{\sum_{s=1}^{T}\exp\left(\boldsymbol{x}_{is}'\boldsymbol{\beta}\right)}\right)
\end{aligned}$$

Concentrated Log Likelihood

- Estimating $\boldsymbol{\beta}$ via maximizing the concentrated likelihood:

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} \log \left(\frac{\exp\left(\boldsymbol{x}_{it}'\boldsymbol{\beta}\right)}{\sum_{s=1}^{T} \exp\left(\boldsymbol{x}_{is}'\boldsymbol{\beta}\right)} \right)$$

  is consistent for fixed T as $N \to \infty$.

- If $\sum_{t=1}^{T} y_{it} = 0$ for any individual, they do not contribute to the log likelihood.
  - $\alpha_i = 0$ in this case, which predicts $y_{it} = 0 \ \forall t$ perfectly for any value of $\boldsymbol{\beta}$.

First-Order Conditions

The concentrated log-likelihood can be rewritten as:

$$\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it} \log\left(\frac{\exp\left(\boldsymbol{x}_{it}'\boldsymbol{\beta}\right)}{\sum_{s=1}^{T}\exp\left(\boldsymbol{x}_{is}'\boldsymbol{\beta}\right)}\right) = \sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}\left[\boldsymbol{x}_{it}'\boldsymbol{\beta} - \log\left(\sum_{s=1}^{T}\exp\left(\boldsymbol{x}_{is}'\boldsymbol{\beta}\right)\right)\right]$$

Differentiating with respect to $\boldsymbol{\beta}$ yields the first-order conditions (recall $\lambda_{it} = \exp\left(\boldsymbol{x}_{it}'\boldsymbol{\beta}\right)$):

$$\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}\left[\boldsymbol{x}_{it} - \frac{1}{\sum_{s=1}^{T}\exp\left(\boldsymbol{x}_{is}'\boldsymbol{\beta}\right)}\left(\sum_{s=1}^{T}\exp\left(\boldsymbol{x}_{is}'\boldsymbol{\beta}\right)\boldsymbol{x}_{is}\right)\right] = \mathbf{0}$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}\left[\boldsymbol{x}_{it} - \frac{\sum_{s=1}^{T}\lambda_{is}\boldsymbol{x}_{is}}{\sum_{s=1}^{T}\lambda_{is}}\right] = \mathbf{0}$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}\left[y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right] = \mathbf{0}$$

Mean-Differencing Transformation

- The first-order condition from the concentrated log likelihood is precisely the sample analogue from a mean-differencing transformation.
- In our model $\mathbb{E}\left[y_{it}|\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{iT}\right] = \alpha_i\lambda_{it} = \alpha_i \exp\left(\boldsymbol{x}_{it}'\boldsymbol{\beta}\right)$.
- Suppose we do the following transformation:

$$\mathbb{E}\left[y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i \middle| \boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{iT}\right] = \alpha_i\lambda_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\alpha_i\bar{\lambda}_i = 0$$

- By the law of iterated expectations $(\mathbb{E}\left[Z\right] = \mathbb{E}\left[\mathbb{E}\left[Z|X\right]\right])$:

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{x}_{it}\left(y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{x}_{it}\left(y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right)\middle| \boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{iT}\right]\right] \\ &= \mathbb{E}\left[\boldsymbol{x}_{it}\underbrace{\mathbb{E}\left[\left(y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right)\middle| \boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{iT}\right]}_{=0}\right] \\ &= \mathbf{0}\end{aligned}$$

Mean-Differencing Transformation

- We can now estimate $\boldsymbol{\beta}$ with the method of moments using the sample analogue:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\left[y_{it}-\frac{\lambda_{it}}{\bar{\lambda}_i}\bar{y}_i\right]=\mathbf{0}$$

  which is exactly the same as the first-order condition for $\boldsymbol{\beta}$ from the concentrated log likelihood.
- $\boldsymbol{\beta}$ is then estimated by finding the $\boldsymbol{\beta}$ that makes the left-hand-side of the above exactly zero.

Poisson Random Effects

- $y_{it}|\alpha_i, \lambda_{it}$ is distributed iid Poisson with rate $\alpha_i \lambda_{it}$, as before.
- However, now we assume that the α_i are iid, and independent of $\boldsymbol{x}_{it}$.
- We can write the likelihood of $y_{i1}, \ldots, y_{iT}$ as:

$$\begin{aligned}\Pr\left(y_{i1}, \ldots, y_{iT}|\lambda_{it}, \boldsymbol{\theta}_\alpha\right) &= \int_0^\infty \Pr\left(y_{i1}, \ldots, y_{iT}|\alpha_i, \lambda_{it}\right) f\left(\alpha_i, \boldsymbol{\theta}_\alpha\right) d\alpha_i \\ &= \int_0^\infty \left[\prod_{t=1}^{T} \Pr\left(y_{it}|\alpha_i, \lambda_{it}\right)\right] f\left(\alpha_i, \boldsymbol{\theta}_\alpha\right) d\alpha_i\end{aligned}$$

- If we specify a distribution for the α_i that is known up to some parameters, we can integrate out the α_i.
- With this we get the distribution of $y_{i1}, \ldots y_{iT}$ conditional on only λ_{it}.

Gamma-Distributed Random Effects

- Suppose we specify

$$f(\alpha_i, \delta) = \frac{\delta^\delta}{\Gamma(\delta)} \alpha_i^{\delta - 1} \exp(-\alpha_i \delta)$$

  which is the density of the Gamma distribution with shape and rate parameter δ

- After many steps (see exercises) we arrive at the likelihood:

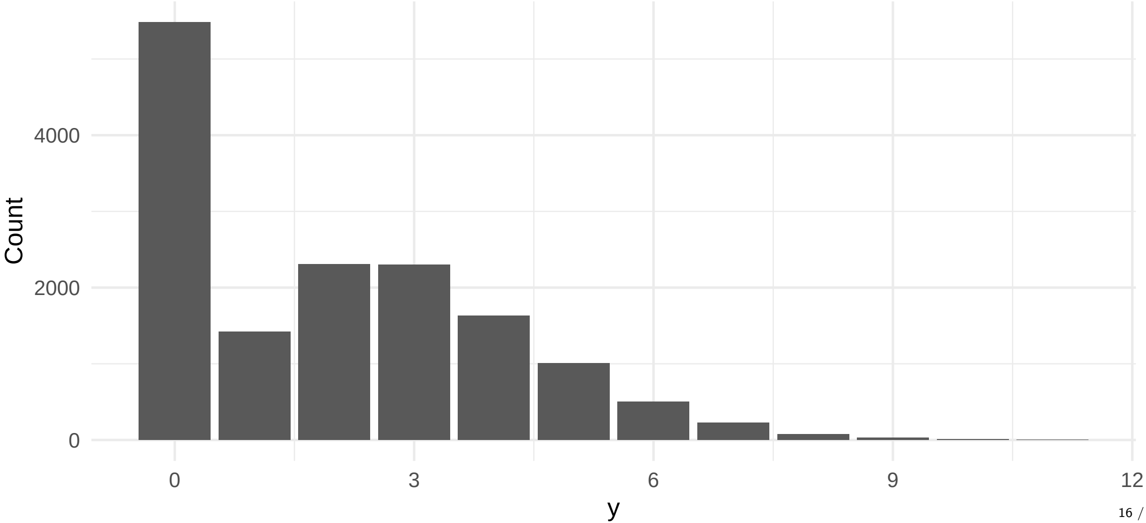
$$\Pr(\boldsymbol{y}_i, \lambda_{it}, \delta) = \left(\prod_{t=1}^{T} \frac{\lambda_{it}^{y_{it}}}{y_{it}!} \right) \frac{\Gamma\left(\sum_{t=1}^{T} y_{it} + \delta \right)}{\Gamma(\delta)} \left(\frac{\delta}{\sum_{t=1}^{T} \lambda_{it} + \delta} \right)^{\delta} \left(\sum_{t=1}^{T} \lambda_{it} + \delta \right)^{-\sum_{t=1}^{T} y_{it}}$$

  which does not depend on α_i.

- We can specify an alternative distribution for the α_i (e.g. log normal), but only the Gamma distribution will result in a closed-form solution (it is a conjugate of the Poisson).

Excess Zeros

- Often your dependent variable may contain more or fewer zeros than would be predicted by the Poisson distribution. For example:



Excess Zeros

- The Hurdle model uses a truncated Poisson for positive y_{it}:

$$\Pr\left(Y_{it} = y_{it}\right) = \begin{cases} \pi & \text{if } y_{it} = 0 \\ (1-\pi)\,\frac{1}{1-e^{-\alpha_i \lambda_{it}}}\,\frac{(\alpha_i \lambda_{it})^{y_{it}} e^{-\alpha_i \lambda_{it}}}{y_{it}!} & \text{if } y_{it} \geq 1 \end{cases}$$

- In the Zero-Inflated Poisson (ZIP) model, $y_{it} \sim 0$ with probability π and $y_{it} \sim \mathit{Poisson}\left(\alpha_i \lambda_{it}\right)$ with probability $1-\pi$, so:

$$\Pr\left(Y_{it} = y_{it}\right) = \begin{cases} \pi + (1-\pi)\, e^{-\alpha_i \lambda_{it}} & \text{if } y_{it} = 0 \\ (1-\pi)\,\frac{(\alpha_i \lambda_{it})^{y_{it}} e^{-\alpha_i \lambda_{it}}}{y_{it}!} & \text{if } y_{it} \geq 1 \end{cases}$$

- In both cases π could be a function of covariates (e.g. logit probabilities).

Dynamic Count Models

- We now introduce the lag of the dependent variable as an explanatory variable.
- There are different possible functional forms.
- One is called the *exponential feedback model*. With one lag it is:

$$y_{it} = \alpha_i \exp\left(\rho y_{it-1} + \boldsymbol{x}'_{it}\boldsymbol{\beta}\right) + u_{it}$$

  - A problem with the expondential feedback model, however, is that if $\rho > 0$, the model can be explosive.

- Another is called the *linear feedback model*. With one lag it is:

$$y_{it} = \rho y_{it-1} + \alpha_i \exp\left(\boldsymbol{x}'_{it}\boldsymbol{\beta}\right) + u_{it}$$

Linear Feedback Model

- The LFM is: $y_{it} = \rho y_{it-1} + \alpha_i \exp\left(\boldsymbol{x}'_{it}\boldsymbol{\beta}\right) + u_{it}$
- We can quasi-difference the α_i out as follows (with $\lambda_{it} = \exp\left(\boldsymbol{x}'_{it}\boldsymbol{\beta}\right)$):

$$
\begin{aligned}
\left(y_{it} - \rho y_{it-1}\right)\frac{\lambda_{it-1}}{\lambda_{it}} - \left(y_{it-1} - \rho y_{it-2}\right) &= \left(\alpha_i\lambda_{it} + u_{it}\right)\frac{\lambda_{it-1}}{\lambda_{it}} - \left(\alpha_i\lambda_{it-1} + u_{it-1}\right) \\
&= \alpha_i\lambda_{it-1} + u_{it}\frac{\lambda_{it-1}}{\lambda_{it}} - \alpha_i\lambda_{it-1} - u_{it-1} \\
&= u_{it}\frac{\lambda_{it-1}}{\lambda_{it}} - u_{it-1}
\end{aligned}
$$

- Assuming $\mathbb{E}\left[u_{it}\frac{\lambda_{it-1}}{\lambda_{it}} - u_{it-1}\middle| y_{i1},\ldots,y_{it-2},\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{it}\right] = 0$, we can use the LHS of the above as moments to estimate ρ and $\boldsymbol{\beta}$.

Linear Feedback Model: Estimation via GMM

- Let:

$$q_{it}(\boldsymbol{\theta}) = (y_{it} - \rho y_{it-1}) \frac{\lambda_{it-1}}{\lambda_{it}} - (y_{it-1} - \rho y_{it-2})$$

  where $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta})'$.

- From the previous slide: $\mathbb{E}\left[q_{it}(\boldsymbol{\theta}) \,|\, y_{it1}, \ldots, y_{it-2}, \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{it}\right] = 0$.
- Let $\boldsymbol{q}_i(\boldsymbol{\theta}) = (q_{i3}(\boldsymbol{\theta}), \ldots, q_{iT}(\boldsymbol{\theta}))'$.
- Let $\boldsymbol{Z}_i$ be a valid instrument matrix for i so $\mathbb{E}\left[\boldsymbol{Z}_i' \boldsymbol{q}_i(\boldsymbol{\theta})\right] = \mathbf{0}$.
- For example, with $T = 4$ with 1 regressor x_{it} we could use:

$$\boldsymbol{Z}_i = \begin{pmatrix} y_{i1} & 0 & 0 & x_{i2} & 0 \\ 0 & y_{i1} & y_{i2} & 0 & x_{i3} \end{pmatrix}$$

Linear Feedback Model: Estimation via GMM

Estimation via GMM is then:

$$\widehat{\boldsymbol{\theta}}_k = \arg\min_{\theta} \left[\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{Z}_i' \boldsymbol{q}_i \left(\boldsymbol{\theta} \right) \right]' \boldsymbol{W}_k^{-1} \left[\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{Z}_i' \boldsymbol{q}_i \left(\boldsymbol{\theta} \right) \right]$$

where $\boldsymbol{W}_1$ could be the identity matrix or:

$$\boldsymbol{W}_1 = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{Z}_i' \boldsymbol{Z}_i$$

and $\boldsymbol{W}_2$ is:

$$\boldsymbol{W}_2 = \frac{1}{N} \sum_{i=1}^{N} \left[\boldsymbol{Z}_i' \boldsymbol{q}_i \left(\widehat{\boldsymbol{\theta}}_1 \right) \right] \left[\boldsymbol{Z}_i' \boldsymbol{q}_i \left(\widehat{\boldsymbol{\theta}}_1 \right) \right]'$$

References/Reading

- In Cameron and Trivedi's *Microeconometrics: Methods and Applications*, Section 23.7 covers count panel data.
- For more detail, Cameron and Trivedi's book *Regression Analysis of Count Data*. Chapter 9 covers panel data.
- Windmeijer, Frank (2006) "GMM for panel count data models."