Ordered Data and Count Data Models

- Ordered data
- Count Data Models
    - The Poisson Regression Model
    - Overdispersion
    - Heterogeneity and the Negative Binomial Regression Model
    - Hurdle and Zero-Inflated Poisson Models.
    - Binomial Regression
    - Models for Panel Data

Ordered data

- In some problems, the variate of interest assumes more than two discrete outcomes, but these are inherently *ordered*.
- Examples that have appeared in the literature include the following: Bond ratings; Results of taste tests; Surveys on the degree of satisfaction with some service; The level of insurance coverage taken by a consumer: none, part, or full; Employment: unemployed, part time, or full time

Ordered data

- Zavoina and McElvey (1975) modelled ordered data using the following latent variable framework:

$$Y_i^* = \mathbf{X}_i'\boldsymbol{\beta}_0 + u_i, \qquad Y_i = \begin{cases} 0 & Y_i^* \leq \mu_0 \\ 1 & \mu_0 < Y_i^* \leq \mu_1 \\ 2 & \mu_1 < Y_i^* \leq \mu_2 \\ \vdots & \vdots \\ J-1 & \mu_{J-1} < Y_i^* \leq \mu_{J-1} \\ J & \mu_{J-1} < Y_i^* \end{cases}$$

  where the *threshold parameters* are such that $0 = \mu_0 < \mu_1 < \cdots < \mu_{J-1}$ and Y_i^* is a latent variable.
- If the distribution of u_i is specified, the *unknown parameters* $\boldsymbol{\beta}$ and $\mu_2, \ldots, \mu_{J-1}$ can be estimated by maximum likelihood.

- Notice that

$$
\begin{aligned}
p_0(\mathbf{X}_i, \boldsymbol{\beta}_0) &= \mathcal{P}\left(Y_i = 0|\mathbf{X}_i\right) = \mathcal{P}\left(\mathbf{X}_i'\boldsymbol{\beta}_0 + u_i \leq 0|\mathbf{X}_i\right) \\
&= \mathcal{P}\left(u_i \leq -\mathbf{X}_i'\boldsymbol{\beta}_0|\mathbf{X}_i\right) \\
p_1(\mathbf{X}_i, \boldsymbol{\beta}_0) &= \mathcal{P}\left(Y_i = 1|\mathbf{X}_i\right) = \mathcal{P}\left(0 < \mathbf{X}_i'\boldsymbol{\beta}_0 + u_i \leq \mu_1|\mathbf{X}_i\right) \\
&= \mathcal{P}\left(u_i \leq \mu_1 - \mathbf{X}_i'\boldsymbol{\beta}_0|\mathbf{X}_i\right) - \mathcal{P}\left(u_i < -\mathbf{X}_i'\boldsymbol{\beta}_0|\mathbf{X}_i\right) \\
&\vdots \\
p_j(\mathbf{X}_i, \boldsymbol{\beta}_0) &= \mathcal{P}\left(Y_i = j|\mathbf{X}_i\right) = \mathcal{P}\left(\mu_{j-1} < \mathbf{X}_i'\boldsymbol{\beta}_0 + u_i \leq \mu_j|\mathbf{X}_i\right) \\
&= \mathcal{P}\left(u_i \leq \mu_j - \mathbf{X}_i'\boldsymbol{\beta}_0|\mathbf{X}_i\right) - \mathcal{P}\left(u_i < \mu_{j-1} - \mathbf{X}_i'\boldsymbol{\beta}_0|\mathbf{X}_i\right) \\
&\vdots \\
p_J(\mathbf{X}_i, \boldsymbol{\beta}_0) &= \mathcal{P}\left(Y_i = J|\mathbf{X}_i\right) = \mathcal{P}\left(\mu_{J-1} < u_i + \mathbf{X}_i'\boldsymbol{\beta}_0|\mathbf{X}_i\right) \\
&= 1 - \mathcal{P}\left(u_i < \mu_{J-1} - \mathbf{X}_i'\boldsymbol{\beta}_0|\mathbf{X}_i\right)
\end{aligned}
$$

Ordered data

- Therefore, the log-likelihood function is simply

$$\log L(\theta) = \sum_{i=1}^{n} \sum_{j=0}^{J} \mathbf{1}\left(Y_i = j\right) \log\left[p_j(\mathbf{X}_i, \boldsymbol{\beta})\right]$$

- As in all discrete choice models, the variance of u_i is *not identified*.
- The *ordered-probit* and *ordered-logit* are the most used special cases of this model.

Ordered data

- For the *ordered-probit*

$$\mathcal{P}\left(u_i \leq \mu_j - \mathbf{X}_i'\boldsymbol{\beta}_0 | \mathbf{X}_i\right) = \Phi\left(\mu_j - \mathbf{X}_i'\boldsymbol{\beta}_0\right)$$

- For the *ordered-logit*

$$\mathcal{P}\left(u_i \leq \mu_j - \mathbf{X}_i'\boldsymbol{\beta}_0 | \mathbf{X}_i\right) = \frac{\exp\left(\mu_j - \mathbf{X}_i'\boldsymbol{\beta}_0\right)}{1 + \exp\left(\mu_j - \mathbf{X}_i'\boldsymbol{\beta}_0\right)}$$

- Interpreting coefficients requires some care. For instance in the *ordered probit* model we have

$$\begin{aligned} \frac{\partial p_0(\mathbf{X}_i, \boldsymbol{\beta}_0)}{\partial x_k} &= -\beta_{0k}\phi(-\mathbf{X}_i'\boldsymbol{\beta}_0), \frac{\partial p_J(\mathbf{X}_i, \boldsymbol{\beta}_0)}{\partial x_k} = \beta_{0k}\phi(\mu_{J-1} - \mathbf{X}_i'\boldsymbol{\beta}_0) \\ \frac{\partial p_j(\mathbf{X}_i, \boldsymbol{\beta}_0)}{\partial x_k} &= \beta_{0k}[\phi(\mu_{j-1} - \mathbf{X}_i'\boldsymbol{\beta}_0) - \phi(\mu_j - \mathbf{X}_i'\boldsymbol{\beta}_0)], j = 1, ..., J-1 \end{aligned}$$

- For $1 < j < J$, the sign of $\partial p_j(\mathbf{X}_i, \boldsymbol{\beta}_0)\partial x_k$ is ambiguous. It depends on $|\mu_{j-1} - \mathbf{X}_i'\boldsymbol{\beta}_0|$ versus $|\mu_j - \mathbf{X}_i'\boldsymbol{\beta}_0|$ (remember, $\phi(\cdot)$ is symmetric about zero).

Ordered data

- The OP and OL models allow us to obtain *sign of the partial effects* on $\mathcal{P}(Y > j|\mathbf{X}_i)$: for a continuous variable x_h. For the OP model

$$\frac{\partial \mathcal{P}(Y_i > j|\mathbf{X}_i)}{\partial x_h} = \beta_h \phi(\mu_j - \mathbf{X}_i' \boldsymbol{\beta}),$$

  If $\beta_h > 0$, an increase in x_h increases the probability that Y_i is greater than any value j.

- Of course the we can interpret the sign of the parameters in the *latent variable model*.

Ordered data

- A closely related model can be used for *grouped data*.
- **Example:** Income reported in non-overlapping intervals
- In this case, the threshold parameters are the limits of the intervals.
- The main difference is that, for $J > 0$, the variance of u_i is *identified* because the thresholds give information on the scale of u_i.

The Poisson Regression Model

- In many relevant applications, the variate of interest is *the count of the number of occurrences of some event in a given period of time* (rare events).
- Examples include: number of accidents, number of patents, number of takeovers, number of purchases, number of doctor visits, number of jobs and number of trips.
- These data have some very specific characteristics:
    - **Discreteness;**
    - **non-negative;**
    - **Many zeros and a long right-hand tail**.
- In this context, standard linear models are *not appealing* because:
    - **The conditional expectation is necessarily non-negative;**
    - **The data is intrinsically heteroskedastic;**
    - **Do not allow the computation of the probability of events of interest.**

The Poisson Regression Model

- The basic model for count data is the *Poisson regression*, defined by

$$\mathcal{P}\left(Y_i = j | \mathbf{X}_i\right) = \frac{\exp\left(-\lambda(\mathbf{X}_i, \boldsymbol{\beta}_0)\right)\lambda(\mathbf{X}_i, \boldsymbol{\beta}_0)^j}{j!}, \qquad j = 0, 1, 2, \ldots$$

$$\mathrm{E}(Y_i | \mathbf{X}_i) = \mathrm{Var}(Y_i | \mathbf{X}_i) = \lambda(\mathbf{X}_i, \boldsymbol{\beta}_0)$$

- Notice, however, that

$$\mathrm{Var}(Y_i) = \mathrm{E}_x\left[\lambda(\mathbf{X}_i, \boldsymbol{\beta}_0)\right] + \mathrm{Var}_x\left[\lambda(\mathbf{X}_i, \boldsymbol{\beta}_0)\right] \geq \mathrm{E}_x\left[\lambda(\mathbf{X}_i, \boldsymbol{\beta}_0)\right] = \mathrm{E}(Y_i).$$

where in general, the following specification is adopted: $\lambda(\mathbf{X}_i, \boldsymbol{\beta}_0) = \exp\left(\mathbf{X}_i'\boldsymbol{\beta}_0\right)$.

- Therefore,

$$\frac{\partial \mathrm{E}(Y_i | \mathbf{X}_i)}{\partial \mathbf{X}_i} = \exp\left(\mathbf{X}_i'\boldsymbol{\beta}_0\right)\boldsymbol{\beta}_0$$

- ML estimation of $\boldsymbol{\beta}_0$ is straightforward.

The Poisson Regression Model

- The log-likelihood function, likelihood equations and the Hessian are given by

$$\log L(\beta) = \sum_{i=1}^{n}\left[-\exp\left(\mathbf{X}_i'\boldsymbol{\beta}\right) + \left(\mathbf{X}_i'\boldsymbol{\beta}\right)Y_i - \log\left(Y_i!\right)\right]$$

$$\frac{\partial \log L\left(\widehat{\boldsymbol{\beta}}\right)}{\partial \beta} = \sum_{i=1}^{n}\left[Y_i - \exp(\mathbf{X}_i'\widehat{\boldsymbol{\beta}})\right]\mathbf{X}_i = 0$$

$$\frac{\partial^2 \log L(\beta)}{\partial\beta\partial\beta'} = -\sum_{i=1}^{n}\exp(\mathbf{X}_i'\boldsymbol{\beta})\mathbf{X}_i\mathbf{X}_i'$$

- Notice that the Hessian is *negative definite* for all $\mathbf{X}$ and $\boldsymbol{\beta}$, which facilitates the estimation and ensures the uniqueness of the maximum, **if it exists**.
- The MLE has the usual properties. In particular

$$\sqrt{n}\left(\hat{\beta}_{ML} - \boldsymbol{\beta}_0\right) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}\left(\exp(\mathbf{X}_i'\boldsymbol{\beta}_0)\mathbf{X}_i\mathbf{X}_i'\right)^{-1}\right)$$

- As usual, inference can be performed using the LR, W and LM tests.

Overdispersion

- The Poisson model imposes (conditional) *equidispersion*, which is very restrictive.
- There are many possible causes for overdispersion:
  - **Measurement error;**
  - **Misspecification of the conditional mean;**
  - **Neglected heterogeneity (random parameter variation).**
- Applied economists tend to focus on the neglected heterogeneity issue, assuming

$$E(Y_i|\mathbf{X}_i, \varepsilon_i) = \exp(\mathbf{X}_i'\boldsymbol{\beta}_0 + \varepsilon_i)$$

$$E\left(\exp\left(\varepsilon_i\right)|\mathbf{X}_i\right) = 1, \qquad \text{Var}\left(\exp\left(\varepsilon_i\right)|\mathbf{X}_i\right) = \sigma^2$$

Overdispersion

- In this particular case

$$E(Y_i|\mathbf{X}_i) = E(\lambda(\mathbf{X}_i, \boldsymbol{\beta}_0)|\mathbf{X}_i) = E_\varepsilon\left[\exp(\mathbf{X}_i'\boldsymbol{\beta}_0 + \varepsilon_i)|\mathbf{X}_i\right] = \exp(\mathbf{X}_i'\boldsymbol{\beta}_0)$$

- Therefore, this sort of neglected heterogeneity does not change the form of the conditional expectation of Y_i.
- Gourieroux, Monfort and Trognon (1984) proved the following *powerful result*: If $E(Y_i|\mathbf{X}_i) = \exp(\mathbf{X}_i'\boldsymbol{\beta}_0)$ is correctly specified and the Likelihood function is constructed using a probability distribution which does not necessarily correspond to the true distribution of the data, but belongs to the *family of linear exponencial distributions*, then the *Quasi-Maximum Likelihood* estimator is consistent for $\boldsymbol{\beta}_0$.

Overdispersion

- The family of linear exponencial distributions includes the *Poisson Distribution*, the *Normal Distribution* (with fixed variance). the *binomial* (with fixed number of trials), the *gamma distribution* (with fixed shape parameter)
- In this particular context the *Quasi-Maximum Likelihood* estimator is sometimes called *Pseudo-Maximum Likelihood Estimator* by some authors.
- Inference is done using the results presented previously for the Quasi-Maximum Likelihood estimator. In particular since the Poisson pseudo-MLE is consistent in presence of this sort of misspecification, valid inference can be based on

$$\sqrt{n}\left(\widehat{\beta}-\beta_0\right)\xrightarrow{d}\mathcal{N}\left(0,A^{-1}BA^{-1}\right)$$

$$A = \mathrm{E}\left[\exp(\mathbf{X}_i'\beta_0)\mathbf{X}_i\mathbf{X}_i'\right] \qquad B=\mathrm{E}\left[\left(y_i-\exp(\mathbf{X}_i'\beta_0)\right)^2\mathbf{X}_i\mathbf{X}_i'\right]$$

Note that

$$\begin{aligned}\text{Var}(Y_i|\mathbf{X}_i) &= \text{E}_\varepsilon\left[\exp(\mathbf{X}_i'\boldsymbol{\beta}_0+\varepsilon_i)\right]+\text{Var}_\varepsilon\left[\exp(\mathbf{X}_i'\boldsymbol{\beta}_0+\varepsilon_i)\right]\\ &= \exp(\mathbf{X}_i'\boldsymbol{\beta}_0)+\sigma^2\exp(2\mathbf{X}_i'\boldsymbol{\beta}_0).\end{aligned}$$

- The presence of *overdispersion* can be tested by testing $H_0:\sigma^2=0$.
- This can be done using the following LM (IM) test statistic (Cox, 1983, and Chesher, 1984)

$$T=\sum_{i=1}^{n}\frac{\left(Y_i-\exp(\mathbf{X}_i'\widehat{\boldsymbol{\beta}})\right)^2-Y_i}{\sqrt{2\sum_{i=1}^{n}\exp(2\mathbf{X}_i'\widehat{\boldsymbol{\beta}})}}\xrightarrow{d}\mathcal{N}(0,1)$$

Overdispersion

- Alternatively, we can regress $\left[\left(Y_i - \exp(\mathbf{X}_i'\widehat{\beta})\right)^2 - Y_i\right] \exp(-\mathbf{X}_i'\widehat{\beta})$ on $\exp(\mathbf{X}_i'\widehat{\beta})$ (or on a constant or other functions of $\exp(\mathbf{X}_i'\widehat{\beta})$) and test the significance of the regressor (Cameron & Trivedi, 1986).
- All these tests can also detect *underdispersion*.
- Overdispersion tests are overplayed in the literature:
  1. in practice, the null is almost always rejected;
  2. if this is the only source of misspecification, the **Poisson pseudo-MLE is still consistent**.
- Other specification tests are available, like the *RESET* test that checks the moment condition

$$\mathrm{E}\left[\left(Y_i - \exp(\mathbf{X}_i'\beta_0)\right)\left(\mathbf{X}_i'\beta_0\right)^2\right] = 0$$

- In practice, the test can be performed by checking the significance of the additional regressor $\left(\mathbf{X}_i'\widehat{\beta}\right)^2$.

Heterogeneity and the Negative Binomial Regression Model

- The assumption that Y_i has a Poisson distribution conditional of $\mathbf{X}_i$ and ε_i with mean $\lambda_i = \exp(\mathbf{X}_i'\boldsymbol{\beta}_0 + \varepsilon_i)$, leads to the compound Poisson regression model

$$\mathcal{P}(Y_i = j|\mathbf{X}_i, \varepsilon_i) = \frac{\exp[-\exp(\mathbf{X}_i'\boldsymbol{\beta}_0 + \varepsilon_i)]\exp(\mathbf{X}_i'\boldsymbol{\beta}_0 + \varepsilon_i)^j}{j!}$$

$$\mathcal{P}(Y_i = j|\mathbf{X}_i) = \int_{-\infty}^{+\infty} \frac{\exp[-\exp(\mathbf{X}_i'\boldsymbol{\beta}_0 + \varepsilon_i)]\exp(\mathbf{X}_i'\boldsymbol{\beta}_0 + \varepsilon_i)^j}{j!} g(\varepsilon_i) d\varepsilon_i$$

  where $g(\varepsilon_i)$ is the density function of ε_i and we assumed that $\mathbf{X}_i$ and ε_i are independent.
- This model can be made operational in different ways:
  1. *Pseudo maximum likelihood* estimation (discussed previously);
  2. *Parametric estimation* for specified $g(\varepsilon_i)$;
  3. *Semiparametric estimation* of $\boldsymbol{\beta}_0$ and $g(\varepsilon_i)$.

Heterogeneity and the Negative Binomial Regression Model

- If $g(\varepsilon_i)$ is specified, the MLE can be obtained, but the estimator *may not be robust* to departures from the additional distributional assumptions.
- Assuming that $\exp(\varepsilon_i) \sim \Gamma\left(\sigma^{-2}, \sigma^2\right)$, $\mathcal{P}(Y_i = j|\mathbf{X}_i)$ is given by the **negative-binomial** (Cameron and Trivedi (1986).denote it as NegBin II) model:

$$\mathcal{P}(Y_i = j|\mathbf{X}_i) = \frac{\Gamma\left(j+\sigma^{-2}\right)\left[1+\sigma^{-2}\exp(-\mathbf{X}_i'\boldsymbol{\beta}_0)\right]^{-j}}{\Gamma\left(\sigma^{-2}\right)\Gamma\left(j+1\right)\left(1+\sigma^2\exp(\mathbf{X}_i'\boldsymbol{\beta}_0)\right)^{\sigma^{-2}}}. \qquad (1)$$

- The Poisson model is obtained as a limiting case when $\sigma^2 \to 0$, but $H_0 : \sigma^2 = 0$ **cannot** be tested with a standard LR or W test.
- If the model (1) is misspecified but $\mathrm{E}(Y_i|\mathbf{X}_i) = \exp(\mathbf{X}_i'\boldsymbol{\beta}_0)$ is correct and σ^{-2} is fixed, the **negative-binomial Psedo-MLE** estimator is consistent for $\boldsymbol{\beta}_0$ This follows from the results of Gourieroux, Monfort and Trognon (1984) and the fact that the *negative-binomial distribution* with σ^{-2} fixed is a member of the family of linear exponencial distributions

Heterogeneity and the Negative Binomial Regression Model

- The score test for $H_0 : \sigma^2 = 0$ is the overdispersion test studied before.
- Other parametric alternatives to the Poisson regression are available.
- A *semiparametric alternative* is to assume that ε has a discrete distribution with Q support points $\alpha_1, \ldots, \alpha_Q$ and corresponding probabilities $\pi_1, \ldots, \pi_Q$, leading to

$$\mathcal{P}(Y_i = j | \mathbf{X}_i) = \sum_{q=1}^{Q} \frac{\exp[-\exp(\mathbf{X}_i' \beta + \alpha_q)] \exp(\mathbf{X}_i' \boldsymbol{\beta}_0 + \alpha_q)^j}{j!} \pi_q,$$

Heterogeneity and the Negative Binomial Regression Model

- **For a given** Q, estimation of $\beta, \alpha_1, \ldots, \alpha_Q$ and $\pi_1, \ldots, \pi_{Q-1}$ can be performed by ML.
- This model can be interpreted as *semiparametric approximation* to a compound Poisson model with unspecified distribution.
- This leads to a consistent estimator if Q is *allowed to increase* at an appropriate rate;
- In practice, the value of Q has to be chosen (for example using an information criterion);
- Inference is complicated by the fact that the number of parameters is not fixed;

Hurdle and Zero-Inflated Poisson Models

- In some cases, the population may be contaminated by individuals for which $Y_i \equiv 0$.
- There are two ways to model this type of data. The *Zero-Inflated Poisson Model* and the *Hurdle Model*
- The *Zero-Inflated Poisson Model*: The zero outcome can arise from one of two regimes.In one regime, the outcome is always zero. In the other, the usual Poisson process is at work
- Let Z_i be a bernoulli random variable such that

$$Z_i = \begin{cases} 0 & \text{with } P\left(Z_i = 0|\mathbf{X}_i\right) = p_i \\ 1 & \text{with } P\left(Z_i = 1|\mathbf{X}_i\right) = 1 - p_i \end{cases}$$

where p_i can be a function of the regressors.

Hurdle and Zero-Inflated Poisson Models

- The observed variable is $Y = ZY^*$ where Y^* is a Poisson random variable independent of Z (conditionally on $\mathbf{X}_i$).
- Let $\mathcal{P}\left(Y_i^* = j|\mathbf{X}_i\right) = \pi_i\left(j;\beta_0\right)$ is the Poisson probability function.
- Note that

$$\begin{aligned}\mathcal{P}(Y_i = 0|\mathbf{X}_i) = \mathcal{P}(Z_i = 0|\mathbf{X}_i) + \mathcal{P}(Z_i = 1|\mathbf{X}_i)\mathcal{P}(Y_i^* = 0|\mathbf{X}_i) \\ = p_i + (1 - p_i)\pi_i\left(0;\beta_0\right)\end{aligned}$$

- Additionally for $j > 0$:

$$\begin{aligned}\mathcal{P}(Y_i &= j|\mathbf{X}_i) = \mathcal{P}(Z_i = 1|\mathbf{X}_i)\mathcal{P}(Y_i^* = j|\mathbf{X}_i) \\ &= (1 - p_i)\pi_i\left(j;\beta_0\right)\end{aligned}$$

- Notice that

$$\begin{aligned}\mathrm{E}\left(Y_i|\mathbf{X}_i\right) &= \sum_{j=0}^{\infty} j\mathcal{P}(Y_i = j|\mathbf{X}_i) = \sum_{j=1}^{\infty} j\mathcal{P}(Y_i = j|\mathbf{X}_i) \\ &= (1 - p_i)E\left(Y_i^*|\mathbf{X}_i\right)\end{aligned}$$

Hurdle and Zero-Inflated Poisson Models

- Therefore the standard pseudo maximum likelihood result does not hold here.
- Then, the log-likelihood function for this *zero-inflated* (Mullahy, 1986) model can bewritten as

$$\log L(\beta) = \sum_{i=1}^{n} \log\{[p_i + (1-p_i)\,\pi_i(0;\beta)]^{\mathbf{1}(Y_i=0)} \times [(1-p_i)\,\pi_i(j;\beta)]^{\mathbf{1}(Y_i>0)}\}$$

Hurdle and Zero-Inflated Poisson Model

- The *Hurdle Model* (Mullahy, 1986): A different extension of the basic count data model is obtained by letting the zero and positive observations be generated by different mechanisms.
- In his formulation, a binary probability model determines whether a zero or a nonzero outcome occurs, then, in the latter case we observe always a positive integer $1, 2, 3, \ldots$
- Consider the Bernoulli random variable

$$W_i = \begin{cases} 1 & \text{with } \mathcal{P}\left(W_i = 1 | \mathbf{X}_i\right) = 1 - q_i \\ 0 & \text{with } \mathcal{P}\left(W_i = 0 | \mathbf{X}_i\right) = q_i \end{cases}$$

where q_i may depend on $\mathbf{X}_i$.

Hurdle and Zero-Inflated Poisson Model

- The observed variable is $Y_i = W_i Y_i^{\star}$, where $Y^{\star}$ can only take values 1,2,3,..., (i.e Pr $\left(Y_i^{\star} = 0|\mathbf{X}_i\right) = 0$) and W_i is conditionally independent of $Y^{\star}$.
- In this case

$$P(Y_i = 0|\mathbf{X}_i) = \mathcal{P}(W_i = 0|\mathbf{X}_i) = q_i$$

- Let $\mathcal{P}\left(Y_i^{\star} = j|\mathbf{X}_i\right) = \pi_i^{\star}\left(j;\beta_0\right)$
- Additionally for $j = 1, 2, ...$

$$\begin{aligned}\mathcal{P}(Y_i &= j|\mathbf{X}_i) = \mathcal{P}(W_i = 1|\mathbf{X}_i)\mathcal{P}(Y_i^{\star} = j|\mathbf{X}_i), \\ &= (1 - q_i)\,\pi_i^{\star}\left(j;\beta_0\right)\end{aligned}$$

- In this case we have

$$\begin{aligned}\mathrm{E}\left(Y_i|\mathbf{X}_i\right) &= \sum_{j=0}^{\infty} j\mathcal{P}(Y_i = j|\mathbf{X}_i) = \sum_{j=1}^{\infty} j\mathcal{P}(Y_i = j|\mathbf{X}_i) \\ &= (1 - q_i)E\left(Y_i^{\star}|\mathbf{X}_i\right)\end{aligned}$$

- Again the standard pseudo maximum likelihood result does not hold here.

Hurdle and Zero-Inflated Poisson Model

- Then, the likelihood function has the form

$$\begin{aligned}\log L(\beta) &= \sum_{i=1}^{n}\{\mathbf{1}(Y_i = 0)(\log q_i) + \mathbf{1}(Y_i > 0)\log(1-q_i) \\ &\quad + \mathbf{1}(Y_i > 0)\log[\pi_i^{\star}(j;\beta)]\}\end{aligned}$$

- Notice that this function is separable.
- Correlated unobserved heterogeneity can be allowed for and integrated-out numerically.

Hurdle and Zero-Inflated Poisson Model

- Usually, $\pi_i^\star(j;\beta_0) = \mathcal{P}(Y_i^\star = j|\mathbf{X}_i)$ is specified as a truncated Poisson of the form

$$\pi_i^\star(j;\beta_0) = \frac{\exp(-\lambda_i)\lambda_i^j}{(1-\exp(-\lambda_i))\,j!}, \qquad j > 0,$$

  with $\lambda_i = \exp\left(\mathbf{X}_i'\beta\right)$.
- However, in this model **there is no real truncation** and therefore an equally validspecification would be

$$\pi_i^\star(j;\beta_0) = \frac{\exp(-\lambda_i)\lambda_i^{j-1}}{(j-1)!}, \qquad j > 0.$$

- When the truncated Poisson specification is used and q_i is specified as

$$q_i = \exp\left(-\exp\left(\mathbf{X}_i'\gamma_0\right)\right),$$

  the null of no hurdle can be tested by testing $H_0 : \boldsymbol{\beta}_0 = \gamma_0$.
- In any case, consistency depends on the distributional assumptions.

Binomial Regression

- Now suppose Y_i is a count variable taking values in $\{0, 1, ..., m_i\}$ for an integer $m_i > 0$. A random draw consists of $(Y_i, m_i, \mathbf{X}_i)$ and, as usual, the sample size is n.
- For example, child mortality within families conditional on number of children ever born m_i. Or, m_i is number of adult children in a family and Y_i is the number of who attended college.
- A natural starting point is to view Y_i as the number of "successes" out of m_i independent Bernoulli (zero-one) trials, with probability of success $0 < p(\mathbf{X}_i, \boldsymbol{\beta}_\mathbf{0}) < 1$. Typically, $p(\mathbf{x}_i, \boldsymbol{\beta}_0) = \Phi(\mathbf{X}_i' \boldsymbol{\beta}_0)$ or $p(\mathbf{X}_i, \boldsymbol{\beta}_0) = \Lambda(\mathbf{X}_i' \boldsymbol{\beta}_0)$.

Binomial Regression

- Under the previous assumptions, Y_i given $(m_i, \mathbf{X}_i)$ has a $Binomial[m_i, p(\mathbf{X}_i, \boldsymbol{\beta}_0)]$ distribution:

$$P(Y_i = j | m_i, \mathbf{X}_i) = \binom{m_i}{j} p(\mathbf{X}_i, \boldsymbol{\beta}_0)^j \left(1 - p(\mathbf{X}_i, \boldsymbol{\beta}_0)\right)^{m_i - j}$$

where $\binom{m_i}{j_i} = \frac{m_i!}{j!(m_i - j)!}$

- The mean and variance are

$$\begin{aligned} \mathrm{E}(Y_i | m_i, \mathbf{X}_i) &= m_i p(\mathbf{X}_i, \boldsymbol{\beta}_0) \\ \mathrm{Var}(Y_i | m_i, \mathbf{X}_i) &= m_i p(\mathbf{X}_i, \boldsymbol{\beta}_0)[1 - p(\mathbf{X}_i, \boldsymbol{\beta}_0)]. \end{aligned}$$

Given standard functional forms for $p(\mathbf{X}_i, \boldsymbol{\beta}_0)$, it is easy to obtain partial effects on the mean.

Binomial Regression

- The Binomial log likelihood is

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{Y_i \log[p(\mathbf{X}_i, \boldsymbol{\beta})] + (m_i - Y_i) \log[1 - p(\mathbf{X}_i, \boldsymbol{\beta})] + \log\{m_i!/[Y_i!(m_i - Y_i)!]\}$$

- MLE estimation is straightforward.
- Importantly, the *Binomial distribution* is in the *linear exponential family*, so only $\mathrm{E}(Y_i|m_i, \mathbf{X}_i)$ needs to be correctly specified to consistently estimate $\boldsymbol{\beta}_0$.

Define $Y_i = (Y_{i1}, \ldots, Y_{iT})'$ and $j_i = (j_{i1}, \ldots, j_{iT})'$, and let

$$\begin{aligned} P(Y_{it} &= j_{it}|\mathbf{X}_{it}, \varepsilon_i) = \frac{\exp\left(-\lambda_{it}\right)\lambda_{it}^{j_{it}}}{j_{it}!} \\ \lambda_{it} &= \exp(\mathbf{X}_{it}'\beta + \varepsilon_i) \\ &= \exp(\mathbf{X}_{it}'\beta)\alpha_i, i = 1, ..., n,\ t = 1, ..., T \end{aligned}$$

where ε_i is a random variable and $\alpha_i = \exp\left(\varepsilon_i\right)$.

- We must assume that $\mathrm{E}(\alpha_i|x_{it})$ is a constant (normalized to 1)
- Based on this assumption we have $\mathrm{E}(y_{it}|x_{it}) = \exp(x'_{it}\beta)$
- The *Poisson Quasi-logLikelhood* is given by (up to additive constants):

$$\log L(\beta) = \sum_{i=1}^{n}\sum_{i=1}^{T}\left[Y_{it}\mathbf{X}'_{it}\beta - \exp(\mathbf{X}'_{it}\beta)\right]$$

- The Poisson Quasi-logLikelhood estimator is consistent under mild assumptions.
- Inference must be based on a robust (to heteroskedasticity and dependence) covariance estimator.
- Inclusion of time dummies in the model is generally recommended.

- We require additional assumptions:
  1. *independence* of the elements of $Y_i = (Y_{i1}, \ldots, Y_{iT})$, conditional on ε_i and $\mathbf{X}_i = (\mathbf{X}_{i1}, ..., \mathbf{X}_{iT})'$;
  2. *strict-exogeneity* of the regressors $\mathrm{E}\left(Y_{it}|\mathbf{X}_{i1}, ..., \mathbf{X}_{iT}, \varepsilon_i\right) = \mathrm{E}\left(Y_{it}|\mathbf{X}_{it}, \varepsilon_i\right)$;
  3. the following distributional assumptions:
     1. $\mathcal{P}(Y_{it} = j_{it}|\mathbf{X}_{it}, \varepsilon_i)$ is given by the *Poisson model*;
     2. distribution of ε_i is *known* and *independent* of $\mathbf{X}_{it}$.

- In this case, $L(\beta) = \prod_{i=1}^{n} L_i(\beta)$, where

$$L_i(\beta) = \int_{-\infty}^{+\infty} \left[\prod_{t=1}^{T} \frac{\exp\left(-\exp(\mathbf{X}_{it}'\beta)\alpha_i\right)\left(\exp(\mathbf{X}_{it}'\beta)\alpha_i\right)^{j_{it}}}{j_{it}!} \right] g(\alpha_i)\, d\alpha$$

- If $\alpha_i = \exp(\varepsilon_i)$ is assumed to have a *gamma distribution*, the model has a closed form based on the negative-binomial distribution.
- Often, it is assumed that α_i has a *log-normal distribution* (no closed form).
- Consistency depends, of course, on the validity of the distributional assumptions.

- There is a consistent *fixed-effects* estimator for the Poisson model, that does not require independence between α_i and the regressors.
- As before, this estimator requires strict-exogeneity and independence of the elements of $Y_i = (Y_{i1}, \ldots, Y_{iT})'$, conditional on ε_i and $\mathbf{X}_i$;
- By the additivity property of the Poisson distribution, we have that
$$\sum_{t=1}^{T} Y_{it} \sim Poisson\left(\sum_{t=1}^{T} \lambda_{it}\right).$$

- It turns out that the distribution of Y_i conditional on $\mathbf{X}_i$, α_i and $\sum_{t=1}^{T} Y_{it}$ does not depend on α_i.
- Indeed, we have (for $j_i = (j_{i1}, ..., j_{iT})'$):

$$\mathcal{P}\left(Y_i = j_i \middle| \mathbf{X}_i, \varepsilon_i, \sum_{t=1}^{T} Y_{it}\right) = \frac{\left(\sum_{t=1}^{T} j_{it}\right)!}{\prod_{t=1}^{T} j_{it}!} \prod_{t=1}^{T} \left(\frac{\exp(\mathbf{X}_{it}'\beta_0)}{\sum_{t=1}^{T} \exp(\mathbf{X}_{it}'\beta_0)}\right)^{j_{it}}.$$

- Write

$$p_t(\mathbf{X}_i, \beta_0) = \frac{\exp(\mathbf{X}_{it}'\beta_0)}{\sum_{t=1}^{T} \exp(\mathbf{X}_{it}'\beta_0)}.$$

- Estimation is simple due to the the fact that the log-likelihood function (up to additive constants) is similar to that of the Conditional Logit model:

$$\log\left(L\left(\beta\right)\right) = \sum_{i=1}^{n}\sum_{t=1}^{T} Y_{it}\log(p_t(\mathbf{X}_i, \beta))$$

- Wooldridge (1999) shows that the estimator **is consistent** even if:
  1. Y_{it} is *not Poisson*.
  2. the elements of $Y_i = (Y_{i1}, \ldots, Y_{iT})$ are *not independent*, conditional on α_i and $\mathbf{X}_{i1}, ..., \mathbf{X}_{iT}$.
- Naturally, if these assumptions do not hold, inference must be based on a *robust* (to heteroskedasticity and dependence) covariance matrix.