A double-hurdle count model for completed fertility data from the developing world

Alfonso Miranda
(alfonso.miranda@cide.edu)
Motivation

- Today it is well recognised that social norms induce special features to completed fertility data.
  - Melkersson y Rooth (2000) suggest that social norms are responsible for the relative excess of 0s and 2s on Swedish fertility data.
  - Santos Silva y Covas (2000) say that, among other reasons, social norms are a factor that make families of an only-child be a relatively rare event in Portugal.

- This creates count data that exhibit underdispersion (i.e. mean > variance).

- Various count data models have been developed to fit well fertility data generated in developed countries.
  - Hurdle count models
  - Zero inflated count models.
Motivation

- Data from developing countries, in contrast, exhibit overdispersion (variance $>\text{mean}$) and do not have an excess of 2s.

- These type of data pose other challenges.
  - An important % of women have many children and move from low to high parities without taking any action to limit their fertility.
  
  - Women with a large family may ‘fall’ into a regime where the opportunity cost of having an extra child is low.
    - Having 3 children may lead to a permanent exit from the labour market. Once out of work, having an extra child carries a relatively small cost.
Hurdle model

First I consider the standard Poisson hurdle model (Mullahy 1986),

\[
P(y_i = j) = \begin{cases} 
\exp (-\mu_{0i}) & \text{si } j = 0 \\
[1 - \exp (-\mu_{0i})] P(y_i | y_i > 0) & \text{en caso contrario},
\end{cases}
\]  

where \( P(y_i | y_i > 0) \) is the conditional probability of \( y_i \) given that a positive count has been observed. In particular \( P(y_i | y_i > 0) \) is a Poisson distribution truncated at 0.

\[
P(y_i = j | y_i > 0) = [1 - \exp (-\mu_{1i})]^{-1} \frac{\exp (-\mu_{1i}) \mu_{1i}^j}{j!}; \quad j = 1, 2, 3, \ldots
\]

\[
\mu_{0i} = \exp (x_{0i}' \beta_0) \\
\mu_{1i} = \exp (x_{1i}' \beta_1)
\]
Double hurdle model

Figure 1. Double-Hurdle Model Structure.
To allow a second hurdle I introduce modifications to $P(y_i|y_i > 0)$.

$$P(y_i = j|y_i > 0) = \begin{cases} 
[1 - \exp(-\mu_{1i})]^{-1} \frac{\exp(-\mu_{1i}) \mu_{1i}^j}{j!} & \text{si } j = 1, 2, 3 \\
1 - \sum_{k=1}^{3} [1 - \exp(-\mu_{1i})]^{-1} \cdot \frac{\exp(-\mu_{1i}) \mu_{1i}^k}{k!} \right] P(y_i|y_i \geq 4), & \text{si } j = 4, 5, 6, \ldots 
\end{cases}$$

with

$$\mu_{1i} = \exp(x_{1i}' \beta_1).$$
The probability of crossing the second hurdle given that the first hurdle was crossed is given by

\[
P (y_i > 3 | y_i > 0) = \left[ 1 - \sum_{k=1}^{3} \left[ 1 - \exp (-\mu_{1i}) \right]^{-1} \frac{\exp (-\mu_{1i}) \mu_{1i}^k}{k!} \right].
\]

To close the model we need to specify a functional form for \( P(y_i | y_i \geq 4) \). For convenience we select a Poisson distribution:

\[
P (y_i | y_i \geq 4) = \left[ 1 - \sum_{k=0}^{3} \frac{\exp (-\mu_{2i}) \mu_{2i}^k}{k!} \right]^{-1} \frac{\exp (-\mu_{2i}) \mu_{2i}^j}{j!} \quad \text{si } j = 4, 5, 6, \ldots
\]

(4)

As usual,

\[
\mu_{2i} = \exp (x_{2i}' \beta_2).
\]
The model is estimated by Maximum likelihood. The likelihood function is given by

\[
L = \prod_{y_i=0} \exp (-\mu_{0i}) \prod_{y_i>0} \left[ 1 - \exp (-\mu_{0i}) \right] \\
\cdot \prod_{1 \leq y_i \leq 3} \left[ 1 - \exp (-\mu_{1i}) \right]^{-1} \frac{\exp (-\mu_{1i}) \mu_{1i}^{y_i}}{y_i!} \\
\cdot \prod_{y_i \geq 4} \left[ 1 - \sum_{k=1}^3 \left[ 1 - \exp (-\mu_{1i}) \right]^{-1} \frac{\exp (-\mu_{1i}) \mu_{1i}^k}{k!} \right] \left[ 1 - \sum_{k=0}^3 \frac{\exp (-\mu_{2i}) \mu_{2i}^k}{k!} \right]^{-1} \frac{\exp (-\mu_{2i}) \mu_{2i}^{y_i}}{y_i!} \tag{5}
\]
. hurdlep fecundidad $myvar, xb1($myvar) xb2($myvar) robust

(Double Hurdle Poisson)

Number of obs = 19477
Wald chi2(9) = 52.95
Log pseudolikelihood = -43980.423
Prob > chi2 = 0.0000

| fecundidad | Coef. | Std. Err. | z    | P>|z|    | [95% Conf. Interval] |
|------------|-------|-----------|------|--------|----------------------|
| xb0        |       |           |      |        |                      |
| catolico   | -0.0525424 | 0.0341243 | -1.54 | 0.124  | -0.1194247 | 0.0143399 |
| lenguaind  | -0.0728384 | 0.0378947 | -1.92 | 0.055  | -0.1471107 | 0.0014339 |
| edu12      | -0.0313502 | 0.048588  | -6.45 | 0.000  | -0.0408733 | -0.021827 |
| c4549      | 0.0229784  | 0.0380156 | 0.60  | 0.546  | -0.0515309 | 0.0974876 |
| c5054      | 0.0493672  | 0.0373468 | 1.32  | 0.186  | -0.0238311 | 0.1225655 |
| c5559      | 0.0224527  | 0.0388909 | 0.58  | 0.564  | -0.053772  | 0.0986774 |
| norte      | 0.055761   | 0.0487709 | 1.14  | 0.253  | -0.0398283 | 0.1513502 |
| centro     | 0.0001234  | 0.0467254 | 0.00  | 0.998  | -0.0914568 | 0.0917035 |
| sur        | 0.0460232  | 0.0520179 | 0.88  | 0.376  | -0.0559301 | 0.1479765 |
| _cons      | 1.154666   | 0.0661379 | 17.46 | 0.000  | 1.025038   | 1.284294  |
The model can be extended to allow unobserved heterogeneity and endogenous fertility change (details in the book).
Conclusions

- Catholic religion is associated with a reduction on the probability of transiting from low to high parities on Mexican fertility data.
  - This result may be explained by a relatively weak opposition by the Catholic church to the use and diffusion of contraceptives in Mexico.
- As expected, women’s education reduces the probability of transiting to counts higher than 3.
The end, thanks!
References


