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

Alfonso Miranda
(alfonso.miranda@cide.edu)

Motivation

- Today its 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 $>$ 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),

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

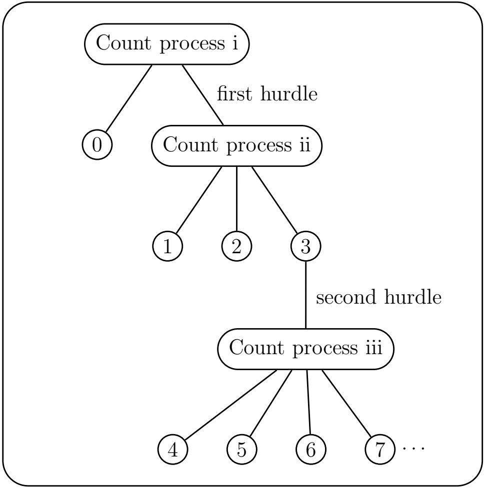
where $\mathrm{P}\left(y_i \mid y_i > 0\right)$ is the conditional probability of y_i given that a positive count has been observed. In particular $\mathrm{P}\left(y_i \mid y_i > 0\right)$ is a Poisson distribution truncated at 0.

$$\mathrm{P}\left(y_i = j \mid y_i > 0\right) = \left[1 - \exp\left(-\mu_{1i}\right)\right]^{-1} \frac{\exp\left(-\mu_{1i}\right)\mu_{1i}^{j}}{j!}; \quad j = 1, 2, 3, \ldots \tag{2}$$

$$\mu_{0i} = \exp\left(\mathbf{x}_{0i}'\boldsymbol{\beta}_0\right)$$
$$\mu_{1i} = \exp\left(\mathbf{x}_{1i}'\boldsymbol{\beta}_1\right)$$

Double hurdle model

Figure 1. Double-Hurdle Model Structure.



To allow a second hurdle I introduce modifications to $\mathrm{P}\left(y_i|y_i>0\right)$.

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

with

$$\mu_{1i}=\exp\left(\mathbf{x}_{1i}'\boldsymbol{\beta}_1\right).$$

The probability of crossing the second hurdle given that the first hurdle was crossed is given by

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

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

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

As usual,

$$\mu_{2i} = \exp\left(\mathbf{x}_{2i}' \boldsymbol{\beta}_2\right).$$

The model is estimated by Maximum likelihood. The likelihood function is given by

$$
\begin{aligned}
L = & \prod_{y_i=0} \exp\left(-\mu_{0i}\right) \prod_{y_i>0} \left[1 - \exp\left(-\mu_{0i}\right)\right] \\
& \cdot \prod_{1\leq y_i\leq 3} \left[1 - \exp\left(-\mu_{1i}\right)\right]^{-1} \frac{\exp\left(-\mu_{1i}\right) \mu_{1i}^{y_i}}{y_i!} \\
& \cdot \prod_{y_i\geq 4} \left[1 - \sum_{k=1}^{3} \left[1 - \exp\left(-\mu_{1i}\right)\right]^{-1} \frac{\exp\left(-\mu_{1i}\right) \mu_{1i}^{k}}{k!}\right] \\
& \cdot \prod_{y_i\geq 4} \left[1 - \sum_{k=0}^{3} \frac{\exp\left(-\mu_{2i}\right) \mu_{2i}^{k}}{k!}\right]^{-1} \frac{\exp\left(-\mu_{2i}\right) \mu_{2i}^{y_i}}{y_i!}
\end{aligned} \tag{5}
$$

```
. hurdlep fecundidad $myvar, xb1($myvar) xb2($myvar) robust
```

(*información suprimida*)

```
Double Hurdle Poisson

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

------------------------------------------------------------------------------
             |               Robust
  fecundidad |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
xb0          |
    catolico |  -.0525424   .0341243    -1.54   0.124    -.1194247    .0143399
   lenguaind |  -.0728384   .0378947    -1.92   0.055    -.1471107    .0014339
       edu12 |  -.0313502   .0048588    -6.45   0.000    -.0408733    -.021827
       c4549 |   .0229784   .0380156     0.60   0.546    -.0515309    .0974876
       c5054 |   .0493672   .0373468     1.32   0.186    -.0238311    .1225655
       c5559 |   .0224527   .0388909     0.58   0.564     -.053772    .0986774
       norte |    .055761   .0487709     1.14   0.253    -.0398283    .1513502
      centro |   .0001234   .0467254     0.00   0.998    -.0914568    .0917035
         sur |   .0460232   .0520179     0.88   0.376    -.0559301    .1479765
       _cons |   1.154666   .0661379    17.46   0.000     1.025038    1.284294
-------------+----------------------------------------------------------------
```

```
xb1          |
    catolico |  -.050893    .015718    -3.24   0.001    -.0816997   -.0200863
   lenguaind |  .0408203   .0193155     2.11   0.035     .0029627     .078678
       edu12 | -.0842011    .002328   -36.17   0.000    -.0887639   -.0796383
       c4549 | -.0535419   .0191292    -2.80   0.005    -.0910344   -.0160495
       c5054 | -.1325749   .0185702    -7.14   0.000    -.1689718    -.096178
       c5559 | -.1769705   .0192836    -9.18   0.000    -.2147656   -.1391754
       norte |  .2523125   .0219407    11.50   0.000     .2093095    .2953156
      centro |  .2616377   .0211541    12.37   0.000     .2201765    .3030989
         sur |  .1597381   .0236188     6.76   0.000     .1134461    .2060302
       _cons |   1.71422   .0314022    54.59   0.000     1.652673    1.775767
-------------+----------------------------------------------------------------
xb2          |
    catolico | -.0347861    .016696    -2.08   0.037    -.0675097   -.0020625
   lenguaind |  .0128803   .0159809     0.81   0.420    -.0184417    .0442023
       edu12 | -.0753265    .002399   -31.40   0.000    -.0800285   -.0706245
       c4549 | -.0911024   .0156799    -5.81   0.000    -.1218344   -.0603704
       c5054 | -.2024501   .0163141   -12.41   0.000    -.2344252    -.170475
       c5559 | -.3029522   .0192311   -15.75   0.000    -.3406444     -.26526
       norte |  .2831148   .0572145     4.95   0.000     .1709765    .3952532
      centro |  .3570204   .0563981     6.33   0.000     .2464822    .4675587
         sur |  .2787026   .0575915     4.84   0.000     .1658253    .3915799
       _cons |  1.775248   .0597473    29.71   0.000     1.658146    1.892351
------------------------------------------------------------------------------
```

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

- Arulampalam, W.; Booth, A. (2001). Learning and Earning: Do Multiple Training Events Pay? A Decade of Evidence from a Cohort of Young British Men, *Economica* 68(271):379-400.
- Cameron, C.; Trivedi, P. (1986). Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests, *Journal of Applied Econometrics*, vol. 1(1):29-53.
- Kohler, Hans-Peter (2000). Fertility decline as a coordination problem, *Journal of Development Economics*, 63(2):231-263.
- Covas, F.; Santos Silva, J.M.C. (2000). A modified hurdle model for completed fertility, *Journal of Population Economics*13(2):173-188.
- Labeaga, JM. (1999). A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco, *Journal of Econometrics* 93(1):49-72.
- Miranda, A. (2013). Un modelo de doble valla para datos de conteo y su aplicación en el estudio de la fecundidad en México. En: *Aplicaciones en Economía y Ciencias Sociales con Stata, Stata Press*.
- Miranda, A. (2010). A double-hurdle count model for completed fertility data from the developing world. DoQSS Working Papers 10-11.
- Melkersson, M. and Rooth, D. (2000). Modeling female fertility using inflated count data models, *Journal of Population Economics*,13(2):189-203.

- Mukesh, E. 2002. The empowerment of women, fertility, and child mortality: Towards a theoretical analysis, *Journal of Population Economics*15(3):433-454.
- Mullahy, J. (1986). Specification and testing of some modified count data models, *Journal of Econometrics* 33(3):341-365.
- Terza, J. (1985). A Tobit-type estimator for the censored Poisson regression model, *Economics Letters* 18(4):361-365.
- Willis, R. (1973). A New Approach to the Economic Theory of Fertility Behavior, *Journal of Political Economy* vol. 81(2):S14-64, Part II.
- Yen, S.; Tang, Chao-Hsiun; Su, Shew-Jiuan. (2001). Demand for traditional medicine in Taiwan: a mixed Gaussian-Poisson model approach, *Health Economics*, 10(3):221-232.
- Yen, S.; Jensen, H. (1996). Determinants of Household Expenditures on Alcohol. Staff General Research Papers 927, Iowa State University, Department of Economics.

- Terza, J. (1985). A Tobit-type estimator for the censored Poisson regression model, *Economics Letters* 18(4):361-365.
- Willis, R. (1973). A New Approach to the Economic Theory of Fertility Behavior, *Journal of Political Economy* vol. 81(2):S14-64, Part II.
- Yen, S.; Tang, Chao-Hsiun; Su, Shew-Jiuan. (2001). Demand for traditional medicine in Taiwan: a mixed Gaussian-Poisson model approach, *Health Economics*, 10(3):221-232.
- Yen, S.; Jensen, H. (1996). Determinants of Household Expenditures on Alcohol. Staff General Research Papers 927, Iowa State University, Department of Economics.