Multiple hurdles models in R: the mhurdle package

Fabrizio Carlevaro[1], Yves Croissant[2,*] , Stéphane Hoareau[3]

1. Département d'économetrie, University of Genève
2. Laboratoire d'économie des transports, Université Lumière Lyon II
3. Universitde la Réunion

* Contact author: yves.croissant@let.ish-lyon.cnrs.fr

Keywords: hurdle models, limited dependent variables, maximum likelihood estimation

In applied econometric studies, the dependent variable often exhibits limited variation, *e.g.*:

- the number of hours of work supplied is non-negative,
- the expenditure in the consumption of a particularly good is non-negative,

In these circumstances, ordinary least squares estimation is biased and inconsistent. However, the model can be estimated consistently using maximum likelihood methods that take into account the censored nature of the dependent variable.

This problem has been treated for a long time in the statistic literature dealing with survival models which are implemented in `R` with the `survival` package. It has also close links with the problem of selection bias, for which some methods are implemented in the `sampleSelection`.

`mhurdle` deals specifically with models where the dependent variable is zero-left censored and may present a large proportion of 0, which is typically the case in household expenditure surveys.

Since the seminal paper of Tobin, a large literature in the econometric field has been developed to deal correctly with this problem of zero observations. More specifically, zero observations may appear for the following three reasons:

- budget constraint: the household would like to consume the good, but his consumer problem has a corner solution because the good is too expensive and/or his income is to low,
- selection: the good is not selected by the household, *i.e.* it's not an argument of its utility function,
- infrequency: the good is bought by the household, but with a low frequency so that zero expenditure may be observed during the survey.

The original *Tobin* takes only the first source of zero into account. With `mhurdle`, the three sources of zero may be introduced in the model.

References

Cragg JG (1971). Some statistical models for limited dependent variables with applications for the demand for durable goods. *Econometrica*, 39(5), pp.829-44.

Deaton A, Irish M (1984). A statistical model for zero expenditures in household budgets. Journal of public economics, 23, pp.59-80.

Tobin J (1958). Estimation of relationships for limited dependent variables. Econometrica, 26, pp.24-36.