

# Package ‘mhurdle’

July 2, 2014

**Version** 1.0-1

**Date** 2013-12-27

**Title** Multiple hurdle Tobit models

**Depends** R (>= 2.10), Formula, truncreg, maxLik, pbivnorm

**Suggests** survival

**Description** Estimation of models with zero left-censored variables.  
Null values may be caused by a selection process, insufficient  
resources or infrequency of purchase.

**License** GPL (>= 2)

**URL** <http://www.r-project.org>

**Author** Yves Croissant [aut, cre],Fabrizio Carlevaro [aut],Stephane Hoareau [aut]

**Maintainer** Yves Croissant <yves.croissant@univ-reunion.fr>

**Repository** CRAN

**Repository/R-Forge/Project** mhurdle

**Repository/R-Forge/Revision** 34

**Repository/R-Forge/DateTimeStamp** 2013-12-27 18:17:53

**Date/Publication** 2013-12-28 08:16:14

**NeedsCompilation** no

## R topics documented:

Comics	2
mhurdle	3
rsq	6
vuongtest	7

<b>Index</b>	<b>8</b>
--------------	----------

---

Comics

*Comics*

---

### Description

a cross section from 2003

*number of observations* : 5159

*observation* : individuals

*country* : France

### Usage

```
data(Comics)
```

### Format

A dataframe containing :

**area** one of rural, small, medium, large and paris,

**income** the income of the household (in thousands of euros per year),

**uc** the number of consumption unit (one for the first two adults, one half for other members of the household),

**size** the number of persons in the household,

**age** the age of the person,

**empl** the kind of job, a factor with 9 levels,

**gender** one of male and female,

**couple** does the person live in couple ? A factor with levels yes and no,

**educ** the number of years of education.

### Source

Web site of the “Institut National de la Statistique et des Etudes Economiques”, <http://www.insee.fr/>

**Description**

mhurdle fits a large set of models relevant when the dependent variable is 0 for a part of the sample.

**Usage**

```

mhurdle(formula, data, subset, weights, na.action,
        start = NULL,
        dist = c("ln", "tn", "n", "bc", "ihs"),
        corr = NULL, ...)
## S3 method for class 'mhurdle'
coef(object,
      which = c("all", "h1", "h2", "h3", "sd", "corr", "tr"), ...)
## S3 method for class 'mhurdle'
vcov(object,
      which = c("all", "h1", "h2", "h3", "sd", "corr", "tr"), ...)
## S3 method for class 'mhurdle'
logLik(object, naive = FALSE, ...)
## S3 method for class 'mhurdle'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)
## S3 method for class 'mhurdle'
summary(object, ...)
## S3 method for class 'summary.mhurdle'
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)

## S3 method for class 'mhurdle'
fitted(object,
      which = c("all", "zero", "positive"), ...)
## S3 method for class 'mhurdle'
predict(object, newdata = NULL, ...)
## S3 method for class 'mhurdle'
update(object, new, ...)

```

**Arguments**

formula	a symbolic description of the model to be fitted,
data	a data.frame,
newdata	a data.frame for which the predictions should be computed,
subset	see <a href="#">lm</a> ,
weights	see <a href="#">lm</a> ,

na.action	see <a href="#">lm</a> ,
start	starting values,
dist	the distribution of the error of the consumption equation: one of "n" (normal), "l" (log-normal) or "t" (truncated normal),
corr	indicates whether the errors of the different equations are correlated. For models with two equations, this can be either "d" for dependent and "i" for independent. For models with three equations, this should be a character of length three containing values of "i" and "d",
naive	a boolean, if TRUE, the likelihood of the naive model is returned,
object, x	an object of class "mhurdle",
new	an updated formula for the update method,
digits	see <a href="#">print</a> ,
width	see <a href="#">print</a> ,
which	which coefficients or covariances should be extracted ? Those of the selection ("h1"), consumption ("h2") or purchase ("h3") equation, the other coefficients "other" (the standard error and the coefficient of corr), the standard error ("sigma") or the coefficient of correlation ("rho"),
...	further arguments.

## Details

mhurdle fits models for which the dependent variable is zero for a part of the sample. Null values of the dependent variable may occur because of one or several mechanisms : good rejection, lack of resources and purchase infrequency. The model is described using a three-parts formula : the first part describes the selection process if any, the second part the regression equation and the third part the purchase infrequency process.  $y \sim 0 \mid x_1 + x_2 \mid z_1 + z_2$  means that there is no selection process.  $y \sim w_1 + w_2 \mid x_1 + x_2 \mid 0$  and  $y \sim w_1 + w_2 \mid x_1 + x_2$  describe the same model with no purchase infrequency process. The second part is mandatory, it explains the positive values of the dependent variable. The dist argument indicates the distribution of the error term. If dist = "n", the error term is normal and (at least part of) the zero observations are also explained by the second part as the result of a corner solution. Several models described in the literature are obtained as special cases :

A model with a formula like  $y \sim 0 \mid x_1 + x_2$  and dist="n" is the Tobit model proposed by Tobin (1958).  $y \sim w_1 + w_2 \mid x_1 + x_2$  and dist="l" or dist="t" is the single hurdle model proposed by Cragg (1971). With dist="n", the double hurdle model also proposed by Cragg (1971) is obtained. With corr="h1" we get the correlated version of this model described by Blundell (1987).

$y \sim 0 \mid x_1 + x_2 \mid z_1 + z_2$  is the P-Tobit model of Deaton and Irish (1984), which can be a single hurdle model if dist="t" or dist="l" or a double hurdle model if dist="n".

## Value

an object of class c("mhurdle", "maxLik").

A "mhurdle" object has the following elements :

coefficients    the vector of coefficients,

vcov	the covariance matrix of the coefficients,
fitted.values	a matrix of fitted.values, the first column being the probability of 0 and the second one the mean values for the positive observations,
logLik	the log-likelihood,
gradient	the gradient at convergence,
model	a data.frame containing the variables used for the estimation,
coef.names	a list containing the names of the coefficients in the selection equation, the regression equation, the infrequency of purchase equation and the other coefficients (the standard deviation of the error term and the coefficient of correlation if <code>corr = TRUE</code> ),
formula	the model formula, an object of class <code>Formula</code> ,
call	the call,
rho	the lagrange multiplier test of no correlation.

## References

Blundell R, Meghir C (1987). Bivariate Alternatives to the Tobit Model. *Journal of Econometrics*, 34, 179-200.

Cragg JG (1971). Some Statistical Models for Limited Dependent Variables with Applications for the Demand for Durable Goods. *Econometrica*, 39(5), 829-44.

Deaton A, Irish M (1984). A Statistical Model for Zero Expenditures in Household Budgets. *Journal of Public Economics*, 23, 59-80.

Tobin J (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26(1), 24-36.

## Examples

```
data("tobin", package = "survival")
# tobit model
model010 <- mhurdle(durable ~ 0 | age + quant | 0, tobin, dist = "n")
# independent double hurdle model
model110i <- mhurdle(durable ~ age | quant | 0, tobin, dist = "n")
# Cragg log-normal single hurdle model
model100il <- mhurdle(durable ~ age | quant | 0, tobin, dist = "ln")
# Cragg truncated-normal single hurdle model
model100it <- update(model100il, dist = "tn")
# a double-hurdle p-tobit
model011i <- mhurdle(durable ~ 0 | quant | age, tobin, dist = "n")
```

---

rsq *R squared and pseudo R squared*

---

### Description

This function computes the R squared for multiple hurdle models. The measure is a pseudo coefficient of determination or may be based on the likelihood.

### Usage

```
rsq(object, type = c("coefdet", "lratio"),  
     adj = FALSE, r2pos = c("rss", "ess", "cor"))
```

### Arguments

object	an object of class "mhurdle",
type	one of "coefdet" or "lratio" to select a pseudo coefficient of correlation or a Mc Fadden like measure based on the likelihood function,
adj	if TRUE a correction for the degrees of freedom is performed,
r2pos	only for pseudo coefficient of determination, should the positive part of the R squared be computed using the residual sum of squares ("rss"), the explained sum of squares ("ess") or the coefficient of correlation between the fitted values and the response (cor).

### Value

a numerical value

### References

McFadden D (1974). The Measurement of Urban Travel Demand. *Journal of Public Economics*, 3, 303-328.

### Examples

```
data("tobin", package = "survival")  
# independent double hurdle model  
model110i <- mhurdle(durable ~ age | quant | 0, tobin, dist = "n")  
rsq(model110i, type = "lratio")  
rsq(model110i, type = "coefdet", r2pos = "rss")
```

---

vuongtest *Vuong test for non-nested models*

---

### Description

The Vuong test is suitable to discriminate between two non-nested models.

### Usage

```
vuongtest(x, y,
          type = c("non-nested", "nested", "overlapping"),
          hyp = FALSE,
          variance = c("centered", "uncentered"),
          matrix = c("large", "reduced")
        )
```

### Arguments

x	a first fitted model of class "mhurdle",
y	a second fitted model of class "mhurdle",
type	the kind of test to be computed,
hyp	a boolean, TRUE if one of the models is assumed to be the true model,
variance	the variance is estimated using the centered or uncentered expression,
matrix	the W matrix can be computed using the general expression large or the reduced matrix reduced (only relevant for the nested case),

### Value

an object of class "htest"

### References

Vuong Q.H. (1989) Likelihood ratio tests for model selection and non-nested hypothesis, *Econometrica*, vol.57(2), pp.307-33.

### See Also

vuong in package pscl.

### Examples

```
data("tobin", package = "survival")
# selection double hurdle model
model110i <- mhurdle(durable ~ age | quant | 0, tobin, dist = "n")
# double-hurdle p-tobit
model011i <- mhurdle(durable ~ 0 | quant | age, tobin, dist = "n")
# Vuong test for strictly non-nested models
vuongtest(model011i, model110i)
```

# Index

## \*Topic **datasets**

Comics, [2](#)

## \*Topic **htest**

rsq, [6](#)

vuongtest, [7](#)

## \*Topic **regression**

mhurdle, [3](#)

coef.mhurdle (mhurdle), [3](#)

Comics, [2](#)

fitted.mhurdle (mhurdle), [3](#)

lm, [3](#), [4](#)

logLik.mhurdle (mhurdle), [3](#)

mhurdle, [3](#)

predict.mhurdle (mhurdle), [3](#)

print, [4](#)

print.mhurdle (mhurdle), [3](#)

print.summary.mhurdle (mhurdle), [3](#)

rsq, [6](#)

summary.mhurdle (mhurdle), [3](#)

update.mhurdle (mhurdle), [3](#)

vcov.mhurdle (mhurdle), [3](#)

vuongtest, [7](#)