Package ‘mhurdle’
September 18, 2018

Version 1.1-8
Date 2018-09-17
Title Multiple Hurdle Tobit Models
Depends R (>= 2.10), methods, Formula, truncreg, maxLik, survival, texreg
Description Estimation of models with zero left-censored variables.
Null values may be caused by a selection process (Cragg (1971) <doi:10.2307/1909582>), insufficient resources
<doi:10.1016/0047-2727(84)90067-7>).
License GPL (>= 2)
URL https://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 75
Repository/R-Forge/DateTimeStamp 2018-09-18 13:53:04
Date/Publication 2018-09-18 17:00:07 UTC
NeedsCompilation yes

R topics documented:

Interview ................................................................. 2
mhurdle ................................................................. 3
rsq ................................................................. 6
vuongtest ................................................................. 7

Index 9
Description

a cross section from 2014

number of observations : 1000

observation : households

country : United-States

Usage

data(Interview)

Format

A dataframe containing :

- **month** the month of the interview,
- **size** the number of person in the household,
- **cu** the number of consumption units in the household,
- **income** the income of the household for the 12 month before the interview,
- **linc** the logarithme of the net income per consumption unit divided by its mean,
- **linc2** the square of link,
- **smsa** does the household live in a SMSA (yes or no),
- **sex** the sex of the reference person of the household (male and female),
- **race** the race of the head of the household, one of white, black, indian, asian, pacific and multirace,
- **hispanic** is the reference person of the household is hispanic (no or yes),
- **educ** the number of year of education of the reference person of the household,
- **age** the age of the reference person of the household - 50,
- **age2** the square of age
- **car** cars in the household,
- **food** food,
- **alcool** ,
- **housing** ,
- **apparel** ,
- **transport** ,
- **health** ,
- **entertainment** ,
perscare, reading, education, tobacco, miscexp, cashcont, insurance, shows, foodaway, vacations.

Source

mhurdle

Estimation of limited dependent variable models

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", "n", "bc", "ihs"), h2 = FALSE, scaled = TRUE, corr = FALSE, robust = TRUE, check.grad = FALSE, ...) ## S3 method for class 'mhurdle'
coef(object, which = c("all", "h1", "h2", "h3", "h4", "sd", "corr", "tr", "pos"), ...) ## S3 method for class 'mhurdle'
vcov(object, which = c("all", "h1", "h2", "h3", "h4", "sd", "corr", "tr", "pos"), ...) ## 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 lm,
- **weights**: see lm,
- **na.action**: see lm,
- **start**: starting values,
- **dist**: the distribution of the error of the consumption equation: one of "n" (normal), "ln" (log-normal) "bc" (box-cox normal) and "ihs" (inverse hyperbolic sinus transformation),
- **h2**: if TRUE the second hurdle is effective, it is not otherwise,
- **scaled**: if TRUE, the dependent variable is divided by its geometric mean,
- **corr**: a boolean indicating whether the errors of the different equations are correlated or not,
- **robust**: transformation of the structural parameters in order to avoid numerical problems,
- **check.grad**: if TRUE, a matrix containing the analytical and the numerical gradient for the starting values are returned,
- **naive**: a boolean, it 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 print,
- **width**: see print,
- **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. \( y \sim 0 \mid x1 + x2 \mid z1 + z2 \) means that there is no selection process. \( y \sim w1 + w2 \mid x1 + x2 \mid 0 \) and \( y \sim w1 + w2 \mid x1 + x2 \) describe the same model with no purchase infrequency process. The second part is mandatory, it explains the positive values of the dependent variable. The \( \text{dist} \) argument indicates the distribution of the error term. If \( \text{dist} = "n" \), 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 \( y \sim 0 \mid x1 + x2 \) and \( \text{dist} = "n" \) is the Tobit model proposed by Tobin (1958).

\( y \sim w1 + w2 \mid x1 + x2 \) and \( \text{dist} = "1" \) or \( \text{dist} = "t" \) is the single hurdle model proposed by Cragg (1971).

With \( \text{dist} = "n" \), the double hurdle model also proposed by Cragg (1971) is obtained. With \( \text{corr} = "h1" \), we get the correlated version of this model described by Blundell (1987).

\( y \sim 0 \mid x1 + x2 \mid z1 + z2 \) is the P-Tobit model of Deaton and Irish (1984), which can be a single hurdle model if \( \text{dist} = "t" \) or \( \text{dist} = "1" \) or a double hurdle model if \( \text{dist} = "n" \).

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 \( \text{corr} = \text{TRUE} \)),
- `formula` the model formula, an object of class `Formula`,
- `call` the call,
- `rho` the lagrange multiplier test of no correlation.

References


Examples

```r
data("Interview", package = "mhurdle")

# independent double hurdle model
idhm <- mhurdle(vacations ~ car + size | linc + linc2 | 0, Interview,
               dist = "ln", h2 = TRUE, method = "bfgs")

# dependent double hurdle model
ddhm <- mhurdle(vacations ~ car + size | linc + linc2 | 0, Interview,
               dist = "ln", h2 = TRUE, method = "bfgs", corr = TRUE)

# a double hurdle p-tobit model
ptm <- mhurdle(vacations ~ 0 | linc + linc2 | car + size, Interview,
               dist = "ln", h2 = TRUE, method = "bfgs", corr = TRUE)
```

---

### 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**

```r
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").
The Vuong test is suitable to discriminate between two non-nested models.

Usage

```r
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

See Also
vuong in package pscl.

Examples

```r
data("Interview", package = "mhurdle")
# dependent double hurdle model
dhm <- mhurdle(vacations ~ car + size | linc + linc2 | 0, Interview,
               dist = "ln", h2 = TRUE, method = "bhhh", corr = TRUE)

# a double hurdle p-tobit model
ptm <- mhurdle(vacations ~ 0 | linc + linc2 | car + size, Interview,
               dist = "ln", h2 = TRUE, method = "bhhh", corr = TRUE)
vuongtest(dhm, ptm)
```
Index

*Topic **datasets**
  Interview, 2
*Topic **htest**
  rsq, 6
  vuongtest, 7
*Topic **regression**
  mhurdle, 3

coef.mhurdle (mhurdle), 3
fitted.mhurdle (mhurdle), 3

Interview, 2

lm, 4
logLik.mhurdle (mhurdle), 3

mhurdle, 3

ndvuongtest (vuongtest), 7

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