Package: mhurdle (via r-universe)

August 17, 2024

Version 1.3-2

Date 2024-06-18

Title Multiple Hurdle Tobit Models

Depends R (>= 2.10)

Imports Formula, truncreg, maxLik, survival, Rdpack, prediction, margins, generics, numDeriv, sandwich, nonnest2, CompQuadForm

Suggests knitr, rmarkdown, lmtest, bookdown, testthat, modelsummary, tibble, broom

Description Estimation of models with dependent variable left-censored at zero. Null values may be caused by a selection process Cragg (1971) <doi:10.2307/1909582>, insufficient resources Tobin (1958) <doi:10.2307/1907382>, or infrequency of purchase Deaton and Irish (1984) <doi:10.1016/0047-2727(84)90067-7>.

VignetteBuilder knitr

NeedsCompilation yes

License GPL (>=2)

Encoding UTF-8

URL https://www.R-project.org

RoxygenNote 7.3.1

Roxygen list(markdown = TRUE)

RdMacros Rdpack

Repository https://blaserlab.r-universe.dev

RemoteUrl https://github.com/ycroissant/mhurdle

RemoteRef HEAD

RemoteSha 78c5dd4e0a7c60d4d51dbc70467fae75cec0f809

Contents

Interview   *Interview*

Description

a cross section from 2014

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 .

Details

number of observations : 1000
observation : households
country : United-States

Source

Consumer Expenditure Survey (CE), program of the US Bureau of Labor Statistics https://www.bls.gov/cex/, interview survey.

---

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_gradient = FALSE,
  ...
)
```

Arguments

formula	a symbolic description of the model to be fitted,
data	a data.frame,
subset	see [stats::lm()](),
weights	see [stats::lm()](),
na.action	see [stats::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_gradient	if TRUE, a matrix containing the analytical and the numerical gradient for the starting values are returned,
...	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 occurs because of one or several mechanisms : good rejection, lack of ressources 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 ~ 0 | x1 + x2 | z1 + z2 means that there is no selection process. y ~ w1 + w2 | x1 + x2 | 0 and y ~ w1 + w2 | x1 + x2 describe the same model with no purchase infrequency process. The second part is mandatory, it explains the positive values of the dependant 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 litterature are obtained as special cases :

A model with a formula like y~0|x1+x2 and dist=″n″ is the Tobit model proposed by (Tobin 1958).

y~w1+w2|x1+x2 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 and Meghir 1987).

y~0|x1+x2|z1+z2 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 corr = TRUE,
- formula: the model formula, an object of class Formula
- 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 AS, 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("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)
```

---

mhurdle.methods   *Methods for mhurdle fitted objects*

Description

specific predict, fitted, coef, vcov, summary, ... for mhurdle objects. In particular, these methods
enables to extract the several parts of the model

Usage

```
## 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'
coef(
  object,
  which = c("all", "h1", "h2", "h3", "sd", "corr", "tr", "pos"),
  ...
)

## 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"), mean = FALSE, ...)

  ## S3 method for class 'mhurdle'
  predict(object, newdata = NULL, what = c("E", "Ep", "p"), ...)

  ## S3 method for class 'mhurdle'
  update(object, new, ...)

  ## S3 method for class 'mhurdle'
  nobs(object, which = c("all", "null", "positive"), ...)

  ## S3 method for class 'mhurdle'
  effects(
    object,
    covariate = NULL,
    data = NULL,
    what = c("E", "Ep", "p"),
    reflevel = NULL,
    mean = FALSE,
    ...
  )
```

Arguments

`object`, `x`	an object of class `"mhurdle"`,
`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.
`naive`	a boolean, it TRUE, the likelihood of the naive model is returned,
`digits`	see [print](#),
`width`	see [print](#),
`mean`	if TRUE, the mean of the effects is returned,
`newdata`, `data`	a `data.frame` for which the predictions or the effectsshould be computed,
`what`	for the `predict` and the `effects` method, the kind of prediction, one of E Ep and p (respectively for expected values in the censored sample, expected values in the truncated sample and probability of positive values),
`new`	an updated formula for the update method,

| covariate | the covariate for which the effect has to be computed, |
| reflevel | for the computation of effects for a factor, the reference level, |

---

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("Interview", package = "mhurdle")
# independent double hurdle model
idhm <- mhurdle(vacations ~ car + size | linc + linc2 | 0, Interview,
              dist = "ln", h2 = TRUE, method = "bfgs")
rsq(idhm, type = "lratio")
rsq(idhm, type = "coefdet", r2pos = "rss")
```

---

vuongtest   *Vuoung 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"),
  true_model = 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,
true_model	a boolean, TRUE if one of the models is asumed 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("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