Package 'EMSS'

January 20, 2025

Type Package

Title Some EM-Type Estimation Methods for the Heckman Selection Model

Version 1.1.1

Author
Kexuan Yang <717260446@qq.com>, Sang Kyu Lee <leesa111@msu.edu>, Jun Zhao <zhaojun2021@hotmail.com>, and H
oung-Moon Kim <hmk966a@gmail.com >

Maintainer Sang Kyu Lee <leesa111@msu.edu>

Description Some EM-type algorithms to estimate parameters for the well-known Heckman selection model are provided in the package. Such algorithms are as follow: ECM(Expectation/Conditional Maximization), ECM(NR)(the Newton-Raphson method is adapted to the ECM) and ECME(Expectation/Conditional Maximization Either). Since the algorithms are based on the EM algorithm, they also have EM's main advantages, namely, stability and ease of implementation. Further details and explanations of the algorithms can be found in Zhao et al. (2020) <doi:10.1016/j.csda.2020.106930>.

Depends R (>= 2.10)

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

Imports sampleSelection, mvtnorm

BugReports https://github.com/SangkyuStat/EMSS/issues

URL https://github.com/SangkyuStat/EMSS

Suggests testthat

NeedsCompilation no

Repository CRAN

Date/Publication 2022-01-10 17:12:42 UTC

Contents

`coef.EMSS` *Getting Coefficients of EM type Sample Selection Model Fits*

Description

`coef` method for a class "EMSS".

Usage

```
## S3 method for class 'EMSS'
coef(object, only = NULL, ...)
```

Arguments

`object`	an object of class "EMSS" made by the function `EMSS`.
`only`	a character value for choosing specific variable's coefficients. Initial value is `NULL`, which shows all variable's coefficients. If "response" is written, only coefficients for response variables will be returned, and if "selection" is written, only coefficients for selection variables will be returned.
`...`	not used, but exists because of the compatibility.

Value

a numeric vector or a list, containing one set or two sets, is given.

Examples

```
# examples continued from EMSS
data(Smoke, package = "EMSS")
ex1 <- EMSS(response = cigs_intervals ~ educ,
           selection = smoker ~ educ + age,
           data = Smoke)
coef(ex1)

data(Smoke, package = "EMSS")
ex2 <- EMSS(response = cigs_intervals ~ educ,
           selection =  smoker ~ educ + age,
           data = Smoke, method="ECMnr")
```

```
coef(ex2)

## example using random numbers with exclusion restriction

N <- 1000
errps <- mvtnorm::rmvnorm(N,c(0,0),matrix(c(1,0.5,0.5,1),2,2) )
xs <- runif(N)
ys <- xs+errps[,1]>0
xo <- runif(N)
yo <- (xo+errps[,2])*(ys>0)

ex3 <- EMSS(response = yo ~ xo,
           selection = ys ~ xs,
           initial.param = c(rep(0,4), 0.3, 0.6), method="ECMnr")
coef(ex3)
```

---

confint.EMSS *Getting Confidence Intervals for Parameters of EM type Sample Selection Model Fits*

Description

confint method for a class "EMSS".

Usage

```
## S3 method for class 'EMSS'
confint(object, parm, level = 0.95, ...)
```

Arguments

object	an object of class "EMSS" made by the function EMSS.
parm	not used, but exists because of the compatibility.
level	a numeric value between 0 and 1 for controlling the significance level of confidence interval; default value is 0.95.
...	not used, but exists because of the compatibility.

Examples

```
# examples continued from EMSS
data(Smoke, package = "EMSS")
ex1 <- EMSS(response = cigs_intervals ~ educ,
           selection = smoker ~ educ + age,
           data = Smoke)
confint(ex1)

data(Smoke, package = "EMSS")
ex2 <- EMSS(response = cigs_intervals ~ educ,
```

```
           selection =  smoker ~ educ + age,
           data = Smoke, method="ECMnr")
confint(ex2)

## example using random numbers with exclusion restriction

N <- 1000
errps <- mvtnorm::rmvnorm(N,c(0,0),matrix(c(1,0.5,0.5,1),2,2) )
xs <- runif(N)
ys <- xs+errps[,1]>0
xo <- runif(N)
yo <- (xo+errps[,2])*(ys>0)

ex3 <- EMSS(response = yo ~ xo,
           selection = ys ~ xs,
           initial.param = c(rep(0,4), 0.3, 0.6), method="ECMnr")
confint(ex3)
```

EMSS *EM type Estimation Methods for the Heckman's Sample Selection Model*

Description

Some algorithms: ECM, ECMnr and ECME can be used to estimate parameters in Heckman selection model and contain the advantages of the EM algorithm: easy implementation and numerical stability. "ECMnr" stands for Expectation/Conditioncal Maximization with Newton-Raphson, and "ECME" for Expectation/Conditional Maximization Either.

Usage

```
EMSS(
  response,
  selection,
  data,
  method = "ECM",
  initial.param = NULL,
  eps = 10^(-10)
)
```

Arguments

response	a formula for the response equation.
selection	a formula for the selection equation.
data	a data frame and data has to be included with the form of data.frame.

method	a character indicating which method to be used. ECM stands for Expectation Conditional Maximization, and ECMnr stands for Expectation Conditioncal Maximization with Newton-Raphson, and ECME for Expectation Conditional Maximization Either.
initial.param	a vector, initial parameter values for the estimation. The length of the initial parameters has to be same as the length of parameters, which are to be estimated.
eps	a numerical error value for the end of the loop. A minimum value that can be arbitrarily set to terminate the iteration of the function, in order to find the optimal parameter estimation.

Details

The dependent variable of the selection equation (specified by argument selection) must have exactly two levels (e.g., 'FALSE' and 'TRUE', or '0' and '1'). The default argument method is "ECM" and the default start values ("NULL") are obtained by two-step estimation of this model through the command selection from the package sampleSelection. NA's are allowed in the data. These are ignored if the corresponding outcome is unobserved, otherwise observations which contain NA (either in selection or outcome) are changed to 0.

Value

ECM returns an object of class "ECM". The object class "ECM" is a list containing the following components.

call	a matched call.
estimate_response	estimated regression coefficients for the response formula.
estimate_selection	estimated regression coefficients for the sample selection formula.
estimate_sigma	an estimated scale paramter for the bivariate normal distribution.
estimate_rho	an estimated correlation coefficient for the bivariate normal distribution.
hessian_mat	hessian matrix for parameters.
resp_leng	the numbers of coefficients for the response formula
select_leng	the numbers of coefficients for the selection formula
Q_value	the vallue of the Q function for EM type algorithms
names_response	names of regression coefficients for the reponse formula.
names_selection	names of regression coefficients for the selection formula.

Background

Heckman selection model is classic to deal with the data where the outcome is partially observed and the missing part is not at random. Heckman (1979) developed 2-step and maximum likelihood estimation (MLE) to do the estimation for this selection model. And these two method are described in R package sampleSelection by Toomet and Henningsen (2008). Zhelonkin et al. (2016) developed robust 2-stage method which performs more robustly than the 2-step method to deal with the

data where outlying observations exist and ssmrob package is available. Zhao et al. (2020) extended EM algorithm to more general cases resulting in three algorithms: ECM, ECM(NR), and ECME. They also own EM algorithm's main advantages, namely, stability and ease of implementation.

References

Heckman, J. (1979) Sample selection bias as a specication error. *Econometrica*, 47, 153-161.

Toomet, O. and Henningsen, A. (2008) Sample selection models in R:Package sampleSelection. *Journal of Statistical Software*, 27, 1-23.

Zhao,J., Kim, H.-J. and Kim, H.-M. (2020) New EM-type algorithms for the Heckman selection model. *Computational Statistics and Data Analysis*, 146, https://doi.org/10.1016/j.csda.2020.106930.

Zhelonkin, M., Genton, M.G. and Ronchetti, E. (2016) Robust inference in sample selection models. *Journal of the Royal Statistical Society Series B*, 78, 805-827.

Examples

```
data(Smoke, package = "EMSS")
ex1 <- EMSS(response = cigs_intervals ~ educ,
           selection = smoker ~ educ + age,
           data = Smoke)
print(ex1)

data(Smoke, package = "EMSS")
ex2 <- EMSS(response = cigs_intervals ~ educ,
           selection =  smoker ~ educ + age,
           data = Smoke, method="ECMnr")
print(ex2)

## example using random numbers with exclusion restriction

N <- 1000
errps <- mvtnorm::rmvnorm(N,c(0,0),matrix(c(1,0.5,0.5,1),2,2) )
xs <- runif(N)
ys <- xs+errps[,1]>0
xo <- runif(N)
yo <- (xo+errps[,2])*(ys>0)
ex3 <- EMSS(response = yo ~ xo,
           selection = ys ~ xs,
           initial.param = c(rep(0,4), 0.3, 0.6), method="ECMnr")
print(ex3)
```

Smoke *Survey Data on Smoking Behaviour*

Description

The Data is the subset of the original data from Mullahy (1985) and Mullahy (1997). The dataset is from Wooldridge (2009) used for researches on cross sectinal data studies. The dataset is also available from Smoke from the package sampleSelection.

Usage

```
data(Smoke, package = "EMSS")
```

Format

a dataframe with 807 observations and 8 variables as below:

educ years of schooling (numeric)

age age of respondents (numeric)

cigpric cigarette price(state), cents per pack (numeric)

income annual income in us dollar (numeric)

restaurn state smoking restrictions for restaurants exist or not (categorical)

smoker smoked at least once or not (categorical)

cigs_intervals number of cigarettes smoked per day, with interval boundaries: 0,5,10,20,50 (numeric)

cigs number of cigarettes smoked per day (numeric)

Source

Wooldridge's dataset is available on https://ideas.repec.org/p/boc/bocins/smoke.html#biblio.

References

Jeffrey, M. Wooldridge (2009) *Introductory Econometrics: A modern approach*, Canada: South-Western Cengage Learning.

Mullahy, John (1985) *Cigarette Smoking: Habits, Health Concerns, and Heterogeneous Unobservables in a Microeconometric Analysis of Consumer Demand*, Ph.D. dissertation, University of Virginia.

Mullahy, John (1997), Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior, *Review of Economics and Statistics*, 79, 596-593.

---

summary.EMSS *Summarizing EM type Sample Selection Model Fits*

Description

summary method for a class "EMSS".

Usage

```
## S3 method for class 'EMSS'
summary(object, tidy = FALSE, conf.int = FALSE, conf.level = 0.95, ...)

## S3 method for class 'summary.EMSS'
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

object	an object of class "EMSS" made by the function EMSS.
tidy	a logical value stands for whether the summary format is in tidy format or not, if TRUE, the summary function will return a tidy format.
conf.int	a logical value stands for whether the confidence interval is included in the tiny format or not. If TRUE, confidence intervals are included. If tidy = FALSE, this parameter does not control anything.
conf.level	a numeric value between 0 and 1 for controlling the significance level of confidence interval; default value is 0.95.
...	not used, but exists because of the compatibility.
x	an object of class "summary.EMSS".
digits	a numeric number of significant digits.

Examples

```
# examples continued from EMSS
data(Smoke, package = "EMSS")
ex1 <- EMSS(response = cigs_intervals ~ educ,
           selection = smoker ~ educ + age,
           data = Smoke)
summary(ex1)

data(Smoke, package = "EMSS")
ex2 <- EMSS(response = cigs_intervals ~ educ,
           selection =  smoker ~ educ + age,
           data = Smoke, method="ECMnr")
summary(ex2)

## example using random numbers with exclusion restriction

N <- 1000
errps <- mvtnorm::rmvnorm(N,c(0,0),matrix(c(1,0.5,0.5,1),2,2) )
xs <- runif(N)
ys <- xs+errps[,1]>0
xo <- runif(N)
yo <- (xo+errps[,2])*(ys>0)

ex3 <- EMSS(response = yo ~ xo,
           selection = ys ~ xs,
           initial.param = c(rep(0,4), 0.3, 0.6), method="ECMnr")
summary(ex3)
```

---

vcov.EMSS *Getting Variance-Covariance Matrix for Parameters of EM type Sample Selection Model Fits*

Description

vcov method for a class "EMSS".

Usage

```
## S3 method for class 'EMSS'
vcov(object, ...)
```

Arguments

object	an object of class "EMSS" made by the function EMSS.
...	not used, but exists because of the compatibility.

Examples

```
# examples continued from EMSS
data(Smoke, package = "EMSS")
ex1 <- EMSS(response = cigs_intervals ~ educ,
           selection = smoker ~ educ + age,
           data = Smoke)
vcov(ex1)

data(Smoke, package = "EMSS")
ex2 <- EMSS(response = cigs_intervals ~ educ,
           selection =  smoker ~ educ + age,
           data = Smoke, method="ECMnr")
vcov(ex2)

## example using random numbers with exclusion restriction

N <- 1000
errps <- mvtnorm::rmvnorm(N,c(0,0),matrix(c(1,0.5,0.5,1),2,2) )
xs <- runif(N)
ys <- xs+errps[,1]>0
xo <- runif(N)
yo <- (xo+errps[,2])*(ys>0)

ex3 <- EMSS(response = yo ~ xo,
           selection = ys ~ xs,
           initial.param = c(rep(0,4), 0.3, 0.6), method="ECMnr")
vcov(ex3)
```

Index