Package 'simex'

May 27, 2009

Type Package

Version 1.4

Date 2009-05-26

Depends stats

Suggests mgcv, nlme

Title SIMEX- and MCSIMEX-Algorithm for measurement error models

Author Wolfgang Lederer <Wolfgang.Lederer@googlemail.com>, Helmut Küchenhoff <Kuechenhoff@stat.uni-muenchen.de>

Maintainer Wolfgang Lederer <Wolfgang.Lederer@googlemail.com>

Description Implementation of the SIMEX-Algorithm by Cook & Stefanski and MCSIMEX by Küchenhoff, Mwalili & Lesaffre

License GPL (>= 2)

Encoding latin1

LazyLoad yes

Repository CRAN

Date/Publication 2009-05-27 08:59:33

R topics documented:

simex-package *Error or misclassification correction in models using (MC)SIMEX*

Description

Package simex is an implementation of the SIMEX–algorithm by Cook and Stephanski and the MCSIMEX–Algorithm by Küchenhof, Mwalili and Lesaffre.

Details

Package:	simex
Type:	Package
Version:	1.4
Date:	2009-04-19
License:	GPL 2 or above
LazyLoad:	yes

The package includes first of all the implementation for the SIMEX– and MCSIMEX–Algorithms. Jackknife and asymptotic variance estimation are implemented. Various methods and analytic tools are provided for a simple and fast access to the SIMEX– and MCSIMEX–Algorithm.

Functions simex() and mcsimex() can be used on models issued from lm(), glm() with asymtotic estimation. Models from nls(), gam() (package **mgcv**), lme() and nlme() (package **nlme**) can also be corrected with these algorithms, but without asymptotic estimations.

Author(s)

Wolfgang Lederer

Maintainer: Wolfgang Lederer,⟨wolfgang.lederer@googlemail.com⟩

References

Lederer, W. and Küchenhoof, H. (2006) A short introduction to the SIMEX and MCSIMEX. *R News*, **6/4**, 26 – 31

See Also

simex, mcsimex, misclass

and for functions generating the initial naive models: lm, glm, nls, gam, lme, nlme

Examples

```
# See example(simex) and example(mcsimex)
```

`diag.block` *Constructs a block diagonal matrix*

Description

The function takes a `list` and constructs a block diagonal matrix with the elements of the list on the diagonal. If `d` is not a list then `d` will be repeated `n` times and written on the diagonal (a wrapper for `kronecker()`)

Usage

```
diag.block(d, n)
```

Arguments

`d`	a `list` of matrices or vectors, or a matrix or vector
`n`	number of repetitions

Value

returns a matrix with the elements of the list or the repetitions of the supplied matrix or vector on the diagonal.

Author(s)

Wolfgang Lederer, ⟨wolfgang.lederer@googlemail.com⟩

See Also

`diag`, `kronecker`

Examples

```
a <- matrix(rep(1, 4), nrow = 2)
b <- matrix(rep(2, 6), nrow = 2)
e <- c(3, 3, 3, 3)
f <- t(e)
d <- list(a, b, e, f)
diag.block(d)
diag.block(a, 3)
```

`mc.matrix` *Build and check misclassification matrices from empirical estimations*

Description

Empirical misclassification matrices to the power of lambda may not exist for small values of lambda. These functions provide methods to estimate the nearest version of the misclassification matrix that satisfies the conditions a misclassification matrix has to fulfill, and to check it (existance for exponents smaller than 1).

Usage

```
build.mc.matrix(mc.matrix, method = "series", tuning = sqrt(.Machine$double.eps),
  diag.cor = FALSE, tol = .Machine$double.eps, max.iter = 100)
check.mc.matrix(mc.matrix, tol = .Machine$double.eps)
```

Arguments

`mc.matrix`	an empirical misclassification matrix
`method`	method used to estimate the generator for the misclassification matrix (see Details)
`tuning`	security parameter for numerical reasons
`diag.cor`	should corrections be substracted from the diagonal or from all values corresponding to the size?
`tol`	tolerance level for series method for convegence
`max.iter`	maximal number of iterations for the serie method to converge

Details

Method "series" constructs a generator via the series

$(Pi - I) - (Pi - I)^2/2 + (Pi - I)^3/3 - \ldots$

Method "log" constructs the generator via taking the log of the misclassification matrix.

Small negative off-diagonal values are corrected and set to (0 + tuning).

The amount used to correct for negative values is added to the diagonal element if `diag.cor = TRUE` and distributed among all values if `diag.cor = FALSE`.

Method "jlt" uses the method described by Jarro et al. (see Israel et al.).

Value

`build.mc.matrix()` returns a misclassification matrix that is the closest estimate for a working misclassification matrix.

`check.mc.matrix()` returns a vector of logicals.

Note

Does not always work! So check properly.

Author(s)

Wolfgang Lederer, ⟨wolfgang.lederer@googlemail.com⟩

References

Israel, R.B., Rosenthal, J.S., Wei, J.Z., Finding generators for Markov Chains via empirical transition matrices, with applications to credit ratings, *Mathematical Finance*, **11**, 245–265

See Also

mcsimex, misclass, diag.block

Examples

```
Pi <- matrix(data = c(0.989, 0.01, 0.001, 0.17, 0.829, 0.001, 0.001, 0.18, 0.819),
    nrow = 3, byrow = FALSE)
check.mc.matrix(list(Pi))
check.mc.matrix(list(build.mc.matrix(Pi)))
build.mc.matrix(Pi)

Pi3 <- matrix(c(0.8, 0.2, 0, 0, 0, 0.8, 0.1, 0.1, 0, 0.1, 0.8, 0.1, 0, 0, 0.3, 0.7),
    nrow = 4)
check.mc.matrix(list(Pi3))
build.mc.matrix(Pi3)
check.mc.matrix(list(build.mc.matrix(Pi3)))

P1 <- matrix(c(1, 0, 0, 1), nrow = 2)
P2 <- matrix(c(0.8, 0.15, 0, 0.2, 0.7, 0.2, 0, 0.15, 0.8), nrow = 3, byrow = TRUE)
P3 <- matrix(c(0.4, 0.6, 0.6, 0.4), nrow = 2)
mc.matrix <- list(P1, P2, P3)
check.mc.matrix(mc.matrix) # TRUE FALSE FALSE
```

---

mcsimex *Misclassification in models using MCSIMEX*

Description

Implementation of the misclassification MCSIMEX algorithm as described by Küchenhoff, Mwalili and Lesaffre.

Usage

```
mcsimex(model, SIMEXvariable, mc.matrix, lambda = c(0.5, 1, 1.5, 2),
  B = 100, fitting.method = "quadratic", jackknife.estimation = "quadratic",
  asymptotic = TRUE)

## S3 method for class 'mcsimex':
print(x, digits = max(3, getOption("digits") - 3), ...)
## S3 method for class 'mcsimex':
summary(object,...)
## S3 method for class 'mcsimex':
plot(x, xlab = expression((1 + lambda)), ylab = colnames(b[, -1]),
  ask = FALSE, show = rep(TRUE, NCOL(b) - 1), ...)
## S3 method for class 'mcsimex':
predict(object, newdata, ...)
## S3 method for class 'mcsimex':
refit(object, fitting.method = "quadratic",
  jackknife.estimation = "quadratic", asymptotic = TRUE, ...)
```

Arguments

`model`	the naive model, the misclassified variable must be a factor
`SIMEXvariable`	vector of names of the variables for which the MCSIMEX-method should be applied
`mc.matrix`	if one variable is misclassified it can be a matrix. If more than one variable is misclassified it must be a list of the misclassification matrices, names must match with the SIMEXvariable names, column- and row-names must match with the factor levels. If a special misclassification is desired, the name of a function can be specified (see details)
`lambda`	vector of exponents for the misclassification matrix (without 0)
`B`	number of iterations for each lambda
`fitting.method`	`linear`, `quadratic` and `loglinear` are implemented (first 4 letters are enough)
`jackknife.estimation`	specifying the extrapolation method for jackknife variance estimation. Can be set to `FALSE` if it should not be performed
`asymptotic`	logical, indicating if asymptotic variance estimation should be done, the option `x = TRUE` must be enabled in the naive model
`x`	object of class 'mcsimex'
`digits`	number of digits to be printed
`object`	object of class 'mcsimex'
`xlab`	optional name for the X-Axis
`ylab`	vector containing the names for the Y-Axis
`ask`	logical. If `TRUE`, the user is asked for input, before a new figure is drawn

show	vector of logicals indicating for wich variables a plot should be produced
newdata	optionally, a data frame in which to look for variables with which to predict. If omitted, the fitted linear predictors are used
...	arguments passed to other functions

Details

if mc.matrix is a function the first argument of that function must be the whole dataset used in the naive model, the second argument must be the exponent (lambda) for the misclassification. The function must return a data.frame containing the misclassified SIMEXvariable. An example can be found below.

Asymptotic variance estimation is only implemented for lm and glm

The loglinear fit has the form g(lambda, GAMMA) = exp(gamma0 + gamma1 * lambda). It is realized via the log() function. To avoid negative values the minimum +1 of the dataset is added and after the prediction later substracted exp(predict(...)) - min(data) - 1.

The 'log2' fit is fitted via the nls() function for direct fitting of the model y ~ exp(gamma.0 + gamma.1 * lambda). As starting values the results of a LS-fit to a linear model with a log transformed response are used. If nls does not converge, the model with the starting values is returned.

refit() refits the object with a different extrapolation function.

Value

An object of class 'mcsimex' which contains:

coefficients	corrected coefficients of the MCSIMEX model,
SIMEX.estimates	the MCSIMEX-estimates of the coefficients for each lambda,
lambda	the values of lambda,
model	the naive model,
mc.matrix	the misclassification matrix,
B	the number of iterations,
extrapolation	the model object of the extrapolation step,
fitting.method	the fitting method used in the extrapolation step,
SIMEXvariable	name of the SIMEXvariables,
call	the function call,
variance.jackknife	the jackknife variance estimates,
extrapolation.variance	the model object of the variance extrapolation,
variance.jackknife.lambda	the data set for the extrapolation,

`variance.asymptotic`
: the asymptotic variance estimates,

`theta`
: all estimated coefficients for each lambda and B,

...

Author(s)

Wolfgang Lederer, ⟨wolfgang.lederer@googlemail.com⟩

References

Küchenhoff, H., Mwalili, S. M. and Lesaffre, E. (2006) A general method for dealing with misclassification in regression: The Misclassification SIMEX. *Biometrics*, **62**, 85 – 96

Küchenhoff, H., Lederer, W. and E. Lesaffre. (2006) Asymptotic Variance Estimation for the Misclassification SIMEX. *Computational Statistics and Data Analysis*, **51**, 6197 – 6211

Lederer, W. and Küchenhoff, H. (2006) A short introduction to the SIMEX and MCSIMEX. *R News*, **6(4)**, 26–31

See Also

`misclass`, `simex`

Examples

```
x <- rnorm(200, 0, 1.142)
z <- rnorm(200, 0, 2)
y <- factor(rbinom(200, 1, (1 / (1 + exp(-1 * (-2 + 1.5 * x -0.5 * z))))))
Pi <- matrix(data = c(0.9, 0.1, 0.3, 0.7), nrow = 2, byrow = FALSE)
dimnames(Pi) <- list(levels(y), levels(y))
ystar <- misclass(data.frame(y), list(y = Pi), k = 1)[, 1]
naive.model <- glm(ystar ~ x + z, family = binomial, x = TRUE, y = TRUE)
true.model  <- glm(y ~ x + z, family = binomial)
simex.model <- mcsimex(naive.model, mc.matrix = Pi, SIMEXvariable = "ystar")

op <- par(mfrow = c(2, 3))
invisible(lapply(simex.model$theta, boxplot, notch = TRUE, outline = FALSE,
    names = c(0.5, 1, 1.5, 2)))
plot(simex.model)

simex.model2 <- refit(simex.model, "line")
plot(simex.model2)
par(op)

# example for a function which can be supplied to the function mcsimex()
# "ystar" is the variable which is to be misclassified
# using the example above
my.misclass <- function (datas, k) {
        ystar <- datas$"ystar"
        p1 <- matrix(data = c(0.75, 0.25, 0.25, 0.75), nrow = 2, byrow = FALSE)
        colnames(p1) <- levels(ystar)
```

```
        rownames(p1) <- levels(ystar)
        p0 <- matrix(data = c(0.8, 0.2, 0.2, 0.8), nrow = 2, byrow = FALSE)
        colnames(p0) <- levels(ystar)
        rownames(p0) <- levels(ystar)
        ystar[datas$x < 0] <-
            misclass(data.frame(ystar = ystar[datas$x < 0]), list(ystar = p1), k = k)[, 1]
        ystar[datas$x > 0] <-
            misclass(data.frame(ystar = ystar[datas$x > 0]), list(ystar = p0), k = k)[, 1]
        ystar <- factor(ystar)
        return(data.frame(ystar))
}

simex.model.differential <- mcsimex(naive.model, mc.matrix = "my.misclass", SIMEXvariable =
```

`misclass` *Generates misclassified data*

Description

Takes a `data.frame` and produces misclassified data. Probabilities for the missclassification are given in the `mc.matrix`.

Usage

```
misclass(data.org, mc.matrix, k)
```

Arguments

`data.org`	`data.frame` containing the factor variables. Must be `factors`
`mc.matrix`	a `list` of matrices giving the probabilities for the misclassification. Names of the `list` must correspond to the variable names in `data.org`. The `colnames` must be named according to the factor levels
`k`	the exponent for the misclassification matrix

Value

A `data.frame` containing the misclassified variables

Author(s)

Wolfgang Lederer, ⟨wolfgang.lederer@googlemail.com⟩

See Also

`mcsimex`, `build.mc.matrix`, `diag.block`

Examples

```
x1 <- factor(rbinom(100, 1, 0.5))
x2 <- factor(rbinom(100, 2, 0.5))

p1 <- matrix(c(1, 0, 0, 1), nrow = 2)
p2 <- matrix(c(0.8, 0.1, 0.1, 0.1, 0.8, 0.1, 0.1, 0.1, 0.8), nrow = 3)

colnames(p1) <- levels(x1)
colnames(p2) <- levels(x2)

x <- data.frame(x1 = x1, x2 = x2)
mc.matrix <- list(x1 = p1, x2 = p2)

x.mc <- misclass(data.org = x, mc.matrix = mc.matrix, k = 1)

identical(x[, 1], x.mc[, 1]) # TRUE
identical(x[, 2], x.mc[, 2]) # FALSE
```

`simex` *Measurement error in models using SIMEX*

Description

Implementation of the SIMEX algorithm for measurement error models according to Cook and Stefanski

Usage

```
simex(model, SIMEXvariable, measurement.error, lambda = c(0.5, 1, 1.5, 2),
  B = 100, fitting.method = "quadratic", jackknife.estimation = "quadratic",
  asymptotic = TRUE)

## S3 method for class 'simex':
print(x, digits = max(3, getOption("digits") - 3), ...)
## S3 method for class 'simex':
summary(object,...)
## S3 method for class 'simex':
plot(x, xlab = expression((1 + lambda)), ylab = colnames(b[, -1]),
  ask = FALSE, show = rep(TRUE, NCOL(b) - 1), ...)
## S3 method for class 'simex':
predict(object, newdata, ...)

refit(object, ...)
## S3 method for class 'simex':
refit(object, fitting.method = "quadratic",
  jackknife.estimation = "quadratic", asymptotic = TRUE, ...)
```

Arguments

model	the naive model
SIMEXvariable	character or vector of characters containing the names of the variables with measurement error
measurement.error	vector of standard deviations of the known measurement errors
lambda	vector of lambdas for which the simulation step should be done (without 0)
B	number of iterations for each lambda
fitting.method	fitting method linear, quadratic, nonlinear are implemented. (first 4 letters are enough)
jackknife.estimation	specifying the extrapolation method for jackknife variance estimation. Can be set to FALSE if it should not be performed
asymptotic	logical, indicating if asymptotic variance estimation should be done, in the naive model the option x = TRUE has to be set
x	object of class 'simex'
digits	number of digits to be printed
object	object of class 'simex'
xlab	optional name for the X-Axis
ylab	vector containing the names for the Y-Axis
ask	logical. If TRUE, the user is asked for input, before a new figure is drawn
show	vector of logicals indicating for wich variables a plot should be produced
newdata	optionally, a data frame in which to look for variables with which to predict. If omitted, the fitted linear predictors are used
...	arguments passed to other functions

Details

Nonlinear is implemented as described in Cook and Stefanski, but is numerically instable. It is not advisable to use this feature. If a nonlinear extrapolation is desired please use the refit() method.

Asymptotic is only implemented for naive models of class lm or glm.

refit() refits the object with a different extrapolation function.

Value

An object of class 'simex' which contains:

coefficients	the corrected coefficients of the SIMEX model,
SIMEX.estimates	the estimates for every lambda,
model	the naive model,

`measurement.error`
: the known error standard deviations,

`B`
: the number of iterations,

`extrapolation`
: the model object of the extrapolation step,

`fitting.method`
: the fitting method used in the extrapolation step,

`residuals`
: the residuals,

`fitted.values`
: the fitted values,

`call`
: the function call,

`variance.jackknife`
: the jackknife variance estimate,

`extrapolation.variance`
: the model object of the variance extrapolation,

`variance.jackknife.lambda`
: the data set for the extrapolation,

`variance.asymptotic`
: the asymptotic variance estimates,

`theta`
: the estimates for every B and lambda,

...

Author(s)

Wolfgang Lederer,⟨wolfgang.lederer@googlemail.com⟩

References

Cook, J.R. and Stefanski, L.A. (1994) Simulation-extrapolation estimation in parametric measurement error models. *Journal of American Statistical Association*, **89**, 1314 – 1328

Carroll, R.J., Küchenhoff, H., Lombard, F. and Stefanski L.A. (1996) Asymptotics for the SIMEX estimator in nonlinear measurement error models. *Journal of the American Statistical Association*, **91**, 242 – 250

Carrol, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C. (2006). *Measurement error in nonlinear models: A modern perspective.*, Second Edition. London: Chapman and Hall.

Lederer, W. and Küchenhoff, H. (2006) A short introduction to the SIMEX and MCSIMEX. *R News*, **6(4)**, 26–31

See Also

`mcsimex` for discrete data with misclassification, `lm`, `glm`

Examples

```
# to test nonlinear extrapolation
#set.seed(3)
x <- rnorm(200, 0, 100)
u <- rnorm(200, 0, 25)
w <- x + u
y <- x + rnorm(200, 0, 9)
true.model <- lm(y ~ x)
naive.model <- lm(y ~ w, x = TRUE)
simex.model <- simex(model = naive.model, SIMEXvariable = "w",
    measurement.error = 25)
plot(x, y)
abline(true.model, col = "darkblue")
abline(simex.model, col = "red")
abline(naive.model, col = "green")
legend(min(x), max(y), legend = c("True Model", "SIMEX model", "Naive Model"),
    col = c("darkblue", "red", "green"), lty = 1)

plot(simex.model, mfrow = c(2, 2))

simex.model2 <-  refit(simex.model, "line")
plot(simex.model2, mfrow = c(2, 2))
```

Index