Package 'heckmanGE'

October 7, 2024

Title Estimation and Inference for Heckman Selection Models with Cluster-Robust Variance

Version 1.0.0

Date 2024-10-02

Description Tools for the estimation of Heckman selection models with robust variance-covariance matrices. It includes functions for computing the bread and meat matrices, as well as clustered standard errors for generalized Heckman models, see Fernando de Souza Bastos and Wagner Barreto-Souza and Marc G. Genton (2022, ISSN: <https://www.jstor.org/stable/27164235>). The package also offers cluster-robust inference with sandwich estimators, and tools for handling issues related to eigenvalues in covariance matrices.

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.2

URL <https://github.com/fsbmat-ufv/heckmanGE>

BugReports <https://github.com/fsbmat-ufv/heckmanGE/issues>

Imports glm2, maxLik, miscTools, vctrs

Depends R (>= 2.10)

LazyData true

NeedsCompilation no

Author Bastos Fernando de Souza [aut, cre]
(<https://orcid.org/0000-0003-1503-4599>),
Barbosa Rogério Jerônimo [aut]
(<https://orcid.org/0000-0002-6796-4547>),
Prates Marcos Oliveira [aut] (<https://orcid.org/0000-0001-8077-4898>)

Maintainer Bastos Fernando de Souza <fernando.bastos@ufv.br>

Repository CRAN

Date/Publication 2024-10-07 12:20:02 UTC

Contents

---

bread.heckmanGE   *bread.heckmanGE*

Description

Bread Function for the `fitheckmanGE` Model

Usage

```
bread.heckmanGE(object, ...)
```

Arguments

object	An object of class `fitheckmanGE`, which is the result of fitting a Heckman selection model using the `fitheckmanGE` function.
...	Additional arguments (currently unused).

Details

This function calculates the "bread" component of the sandwich estimator for the `fitheckmanGE` model. The bread matrix is typically defined as the product of the number of observations and the variance-covariance matrix of the estimated parameters.

The bread matrix is an essential component of the sandwich estimator used to obtain robust standard errors. It reflects the variability in the estimated parameters due to the model's fit. The function uses the number of observations and the variance-covariance matrix from the `fitheckmanGE` model object to compute this matrix.

Value

A matrix representing the bread component of the sandwich estimator. The matrix is calculated as the product of the number of observations and the variance-covariance matrix of the estimated parameters.

`coef.heckmanGE`	*Extract Coefficients from a Generalized Heckman Model*

Description

This function extracts the coefficients from a `heckmanGE` class object. You can specify which parts of the model you want to retrieve the coefficients for: selection, outcome, dispersion, or correlation. By default, the function returns the complete coefficient vector for all parts.

Usage

```
## S3 method for class 'heckmanGE'
coef(
  object,
  part = c("selection", "outcome", "dispersion", "correlation"),
  ...
)
```

Arguments

`object`	An object of class `heckmanGE`, which contains the fitted model.
`part`	A character vector indicating which parts of the model coefficients to return. Valid options are: `"selection"`, `"outcome"`, `"dispersion"`, and `"correlation"`. Multiple parts can be specified. By default, all parts are included in the returned coefficients.
`...`	Additional arguments passed to or from other methods. Currently, these are not used in this method but must be included to match the generic method signature.

Details

The `coef.heckmanGE` function retrieves coefficients from the `heckmanGE` model object based on the specified parts. The parts represent different components of the Heckman model:

- `"selection"`: Coefficients related to the selection equation.
- `"outcome"`: Coefficients related to the outcome equation.
- `"dispersion"`: Coefficients related to the dispersion equation.
- `"correlation"`: Coefficients related to the correlation between selection and outcome.

By default, the function returns coefficients from all parts. You can specify one or more parts in the `part` argument to extract coefficients from specific components.

Value

A numeric vector containing the coefficients extracted from the model object. The coefficients correspond to the specified model parts, returned in the order they are requested.

Examples

```
data(MEPS2001)
selectEq  <- dambexp ~ age + female + educ + blhisp + totchr + ins + income
outcomeEq <- lnambx ~ age + female + educ + blhisp + totchr + ins
dispersion  <- ~ age + female + totchr + ins
correlation  <- ~ age
fit <- heckmanGE(selection = selectEq,
                 outcome = outcomeEq,
                 dispersion = dispersion,
                 correlation = correlation,
                 data = MEPS2001)
# Extracting all coefficients:
coef(fit)

# Extracting only the selection and outcome coefficients:
coef(fit, part = c("selection", "outcome"))
```

---

estfun.heckmanGE   *Compute Estimating Functions for Generalized Heckman Model*

Description

This function calculates the estimating functions (i.e., the gradient of the log-likelihood) for the Generalized Heckman model. It is primarily used for model diagnostics and inference, providing the gradient for each observation with respect to model parameters.

Usage

```
estfun.heckmanGE(x, ...)
```

Arguments

x
: An object of class `heckmanGE`, typically the result of fitting a generalized Heckman model. This object should contain model responses, model matrices, weights, and coefficient indexes necessary for the computation of the gradient.

...
: Additional arguments (currently not used, reserved for future extensions).

Details

The function computes the gradient of the log-likelihood function for the Generalized Heckman model, which includes the selection, outcome, dispersion, and correlation components.

The gradient is calculated per observation, and internally, the helper function `gradlik_gen_i` computes the gradient for each observation given the model parameters. This involves extracting components such as model matrices, weights, and coefficient indexes, and performing matrix operations specific to the model's structure.

Value

A matrix of dimensions n x p, where n is the number of observations and p is the number of parameters in the model. Each element of the matrix corresponds to the gradient of the log-likelihood function with respect to a given parameter for each observation.

Examples

```
# Assuming 'model' is a fitted object of class 'heckmanGE':
data(MEPS2001)
selectEq  <- dambexp ~ age + female + educ + blhisp + totchr + ins + income
outcomeEq <- lnambx ~ age + female + educ + blhisp + totchr + ins
dispersion  <- ~ age + female + totchr + ins
correlation  <- ~ age
fit <- heckmanGE(selection = selectEq,
                 outcome = outcomeEq,
                 dispersion = dispersion,
                 correlation = correlation,
                 data = MEPS2001)
estfun.heckmanGE(fit)
```

---

fitheckmanGE   *fitheckmanGE*

Description

Newton-Raphson Optimization for Generalized Heckman Model Estimation

Usage

```
fitheckmanGE(start, YS, XS, YO, XO, Msigma, Mrho, w)
```

Arguments

start	A numeric vector of initial parameter guesses for the selection, outcome, dispersion, and correlation equations.
YS	A binary vector indicating selection status (1 if selected, 0 otherwise).
XS	A matrix of independent variables for the selection equation.
YO	A numeric vector of observed outcomes (dependent variable) for the outcome equation.
XO	A matrix of independent variables for the outcome equation.
Msigma	A matrix representing the predictors for the dispersion parameter.
Mrho	A matrix representing the predictors for the correlation parameter.
w	A numeric vector of observation weights, used in the likelihood computation.

Details

This function estimates the parameters of a generalized Heckman selection model using a Newton-Raphson optimization algorithm. It supports the modeling of selection and outcome equations, along with associated dispersion and correlation structures.

This function uses the Newton-Raphson algorithm to estimate the parameters of a generalized Heckman model, which accounts for sample selection bias. The model is composed of a selection equation (modeled by YS and XS), an outcome equation (modeled by YO and XO), and additional equations for dispersion (Msigma) and correlation (Mrho). The optimization process maximizes the log-likelihood of the model, allowing for robust estimation of selection bias, while also estimating associated dispersion and correlation parameters.

The function outputs the coefficients, fitted values, residuals, and several information criteria for model comparison.

Value

A list with the following components:

coefficients Named vector of estimated coefficients for selection, outcome, dispersion, and correlation equations.

fitted.values Named list with fitted values for each equation (selection, outcome, dispersion, correlation).

residuals Numeric vector of residuals for the selection and outcome equations.

loglik Log-likelihood value of the fitted model.

vcov Variance-covariance matrix of the estimated parameters.

aic Akaike Information Criterion (AIC) for the model.

bic Bayesian Information Criterion (BIC) for the model.

optimization Details of the optimization process, including convergence information.

---

fitted.heckmanGE *fitted.heckmanGE Extract Fitted Values of the Generalized Heckman Model*

Description

This function extracts the fitted values from a heckmanGE object. You can specify which part of the model you want to retrieve the fitted values for: selection, outcome, dispersion, or correlation. By default, it returns the fitted values for the outcome part of the model.

Usage

```
## S3 method for class 'heckmanGE'
fitted(
  object,
  part = c("selection", "outcome", "dispersion", "correlation"),
  ...
)
```

Arguments

object	An object of class heckmanGE. This object should contain the fitted values for different parts of the model.
part	A character vector specifying which part of the model to return the fitted values for. Options are "selection", "outcome" (default), "dispersion", or "correlation". If multiple parts are provided, only the "outcome" part will be returned.
...	Additional arguments passed to or from other methods. These are not used in this method but must be included to match the generic method signature.

Details

- If part is "selection", the function returns the fitted values from the selection part of the model.
- If part is "outcome", the function returns the fitted values from the outcome part of the model.
- If part is "dispersion", the function returns the fitted values from the dispersion part of the model.
- If part is "correlation", the function returns the fitted values from the correlation part of the model.

If part is not one of the specified options, an error will be raised. If multiple parts are provided, the function defaults to returning the fitted values for the outcome part of the model.

Value

A vector of fitted values corresponding to the specified part of the heckmanGE model. The type of the returned values depends on the part specified.

| heckmanGE | *heckmanGE: A Package for Fitting Sample Selection Models* |

Description

This package provides functions for fitting sample selection models, specifically the Heckman-Ge model. It includes functionality for specifying selection and outcome equations, as well as adjusting parameters for dispersion and correlation.

Estimates the parameters of the Generalized Heckman model

Usage

```
heckmanGE(
  selection,
  outcome,
  dispersion,
  correlation,
  data = sys.frame(sys.parent()),
  weights = NULL,
  cluster = NULL,
  start = NULL
)
```

Arguments

selection	A formula. Selection equation.
outcome	A formula. Outcome Equation.
dispersion	A right-handed formula. The equation for fitting of the Dispersion Parameter.
correlation	A right-handed formula. The equation for fitting of the Correlation Parameter.
data	A data.frame.
weights	an optional vector of weights to be used in the fitting process. Should be NULL or a numeric vector.
cluster	a variable indicating the clustering of observations, a list (or data.frame) thereof, or a formula specifying which variables from the fitted model should be used. See documentation for sandwich::vcovCL. A formula or list specifying the clusters for robust standard errors. Clustering adjusts the standard errors by accounting for correlations within clusters.
start	Optional. A numeric vector with the initial values for the parameters.

Details

Optimized Function for fitting the Generalized Heckman Model

(original version: package ssmodels. Modified by Rogerio Barbosa)

The heckmanGE() function fits a generalization of the Heckman sample selection model, allowing sample selection bias and dispersion parameters to depend on covariates. The `heckmanGE()` function fits a generalization of the Heckman sample selection model, and is compatible with robust variance-covariance estimation using packages such as **sandwich**. In particular, the [vcovCL](#) function can be used for clustering, which adjusts the standard errors by accounting for intra-cluster correlations in the data.

Value

A list of results from the fitted model, including parameter estimates, the Hessian matrix, number of observations, and other relevant statistics. If initial values are not provided, the function estimates them using the Heckman two-step method.

A list containing:

call The matched function call.

coefficients Estimated coefficients for the selection, outcome, dispersion, and correlation equations.

vcov The covariance matrix of the estimated coefficients.

logLik The log-likelihood of the fitted model.

model.frames List of model frames for each equation (selection, outcome, dispersion, and correlation).

fitted.values Fitted values of the outcome equation.

Author(s)

Fernando de Souza Bastos

See Also

[vcovCL](#) for computing robust standard errors with clustering. The function is compatible with the **sandwich** package for estimating heteroskedasticity-consistent and cluster-robust standard errors. This can be useful for adjusting the standard errors when dealing with grouped or clustered data. For more details, see the documentation for [vcovCL](#).

Examples

```
data(MEPS2001)
selectEq  <- dambexp ~ age + female + educ + blhisp + totchr + ins + income
outcomeEq <- lnambx ~ age + female + educ + blhisp + totchr + ins
dispersion  <- ~ age + female + totchr + ins
correlation  <- ~ age
fit <- heckmanGE(selection = selectEq,
                 outcome = outcomeEq,
                 dispersion = dispersion,
                 correlation = correlation,
                 data = MEPS2001)
summary(fit)
```

meat.heckmanGE   *meat.heckmanGE Compute Meat of the Covariance Matrix for the Generalized Heckman Model*

Description

This function calculates the "meat" of the covariance matrix for a heckmanGE model. The "meat" refers to the part of the covariance matrix that is not accounted for by the model's fixed components. This is typically used in conjunction with the "bread" component to form a robust covariance matrix estimator.

Usage

```
meat.heckmanGE(x, adjust = FALSE, ...)
```

Arguments

x   An object of class heckmanGE. This object should be a fitted model for which the covariance matrix is to be computed.

adjust   A logical value indicating whether to apply a small-sample correction to the covariance matrix. If TRUE, the matrix is adjusted by multiplying it by n / (n - k), where n is the number of observations and k is the number of parameters.

...   Additional arguments passed to estfun.heckmanGE, which computes the estimating functions used in the calculation.

Details

- The function calculates the covariance matrix based on the estimating functions obtained from estfun.heckmanGE.

- The "meat" is calculated as the cross-product of the estimating functions, divided by the number of observations. If adjust is TRUE, a small-sample correction is applied.

Value

A matrix representing the "meat" of the covariance matrix. The dimensions and row/column names of the matrix correspond to the number of parameters in the model.

---

meatCL.heckmanGE   *meatCL.heckmanGE Compute the Meat Matrix for a Heckman-Ge Model with Clustering*

Description

This function calculates the meat matrix for a Heckman-Ge model, which is used in the context of clustered standard errors. The meat matrix represents the variability of the estimated parameters and is a crucial component for robust inference.

Usage

```
meatCL.heckmanGE(
  x,
  cluster = NULL,
  type = NULL,
  cadjust = TRUE,
  multi0 = FALSE,
  ...
)
```

Arguments

x	An object of class heckmanGE containing the results from a Heckman-Ge model fit.
cluster	A vector or a data frame specifying the cluster variable(s). If NULL, the default clustering variable is used.
type	The type of heteroscedasticity-consistent (HC) estimator to use. Options are "HC0", "HC1", "HC2", or "HC3". Defaults to "HC0".
cadjust	A logical value indicating whether to adjust for the number of clusters. Defaults to TRUE.
multi0	A logical value indicating whether to include a column of ones in the cluster variable matrix. Defaults to FALSE.
...	Additional arguments passed to other methods.

Value

A matrix representing the meat component of the robust covariance matrix estimator for the Heckman-Ge model.

MEPS2001 *Medical Expenditure Panel Survey (MEPS) Data*

Description

The MEPS dataset contains large-scale survey data from the United States, focusing on health services usage, costs, and insurance coverage. This dataset is restricted to individuals aged 21 to 64 years. It includes outpatient cost data with some zero expenditure values for model adjustment.

Usage

    MEPS2001

Format

A data frame with 3328 observations on the following variables:

- educ: Education status (numeric)
- age: Age (numeric)
- income: Income (numeric)
- female: Gender (binary)
- vgood: Self-reported health status, very good (numeric)
- good: Self-reported health status, good (numeric)
- hospexp: Hospital expenditures (numeric)
- totchr: Total number of chronic diseases (numeric)
- ffs: Family support (numeric)
- dhospexp: Dummy variable for hospital expenditures (binary)
- age2: Age squared (numeric)
- agefem: Interaction between age and gender (numeric)
- fairpoor: Self-reported health status, fair or poor (numeric)
- year01: Year of survey (numeric)
- instype: Type of insurance (numeric)
- ambexp: Ambulatory expenditures (numeric)
- lambexp: Log of ambulatory expenditures (numeric)
- blhisp: Ethnicity (binary)
- instype_s1: Insurance type, version 1 (numeric)
- dambexp: Dummy variable for ambulatory expenditures (binary)
- lnambx: Log-transformed ambulatory expenditures (numeric)
- ins: Insurance status (binary)

Source

2001 Medical Expenditure Panel Survey by the Agency for Healthcare Research and Quality.

Examples

```
data(MEPS2001)
selectEq  <- dambexp ~ age + female + educ + blhisp + totchr + ins + income
outcomeEq <- lnambx ~ age + female + educ + blhisp + totchr + ins
dispersion  <- ~ age + female + totchr + ins
correlation  <- ~ age
fit <- heckmanGE(selection = selectEq,
                 outcome = outcomeEq,
                 dispersion = dispersion,
                 correlation = correlation,
                 data = MEPS2001)
summary(fit)
```

---

model.frame.heckmanGE *model.frame.heckmanGE Get the Model Frames of a Generalized Heckman Regression*

Description

Extracts the model frames for different parts of a heckmanGE model. The model frames include the data used in the regression analysis for each component of the Generalized Heckman Model.

Usage

```
## S3 method for class 'heckmanGE'
model.frame(
  formula,
  part = c("selection", "outcome", "dispersion", "correlation"),
  ...
)
```

Arguments

formula	An object of class heckmanGE. This object should be a fitted model from which the model frames will be extracted.
part	A character vector specifying the model part for which to extract the model frame. Options include "selection", "outcome", "dispersion", and "correlation". The default is "outcome". If multiple parts are specified, only the "outcome" part will be returned.
...	Additional arguments passed to or from other methods. These are not used in this method but must be included to match the generic method signature.

Details

- The function extracts the model frame corresponding to the specified part of the `heckmanGE` model.
- If the `part` argument is not specified correctly or includes multiple parts, the function defaults to returning the model frame for the "outcome" part.

Value

A model frame for the specified part of the `heckmanGE` object. If `part` is not one of the valid options, an error is raised.

---

model.matrix.heckmanGE

model.matrix.heckmanGE Get the Design Matrices of a Generalized Heckman Regression

---

Description

Extracts the design matrices for different parts of a `heckmanGE` model. The design matrices include the predictors used in the regression analysis for each component of the Generalized Heckman Model.

Usage

```
## S3 method for class 'heckmanGE'
model.matrix(
  object,
  part = c("selection", "outcome", "dispersion", "correlation"),
  ...
)
```

Arguments

object	An object of class `heckmanGE`. This object should be a fitted model from which the design matrices will be extracted.
part	A character vector specifying the model part for which to extract the design matrix. Options include "selection", "outcome", "dispersion", and "correlation". The default is "outcome". If multiple parts are specified, only the "outcome" part will be returned.
...	Additional arguments passed to or from other methods. These are not used in this method but must be included to match the generic method signature.

Details

- The function extracts the design matrix corresponding to the specified part of the `heckmanGE` model.
- If the `part` argument is not specified correctly or includes multiple parts, the function defaults to returning the design matrix for the "outcome" part.

Value

A design matrix for the specified part of the `heckmanGE` object. If `part` is not one of the valid options, an error is raised.

---

pnadC_y2024q2 *PNAD Continua de 2024, 2 trimestre*

Description

The Continuous National Household Sample Survey (PNAD Continua) for the second quarter of 2024 is an important source of statistical data in Brazil, conducted by the Brazilian Institute of Geography and Statistics (IBGE). The survey aims to provide up-to-date information on the socioe-conomic characteristics of the Brazilian population, covering topics such as employment, income, education, and other crucial aspects for the formulation of public policies and economic and social studies.

Usage

pnadC_y2024q2

Format

A data frame with 326018 observations on the following variables:

- PSU: Primary Sampling Unit identifier (factor)

- weight: Survey weight (numeric)

- age: Age of the respondent (numeric)

- participation: Labor force participation status (factor)

- male: Male indicator (binary)

- white: White indicator (binary)

- hhold_head: Household head indicator (binary)

- hhold_spouse: Spouse of household head indicator (binary)

- yearsSchooling: Total years of schooling completed (numeric)

- classWorker_employer: Employer indicator (binary)

- classWorker_selfEmployed: Self-employed indicator (binary)

- ln_salary: Natural logarithm of salary (numeric)

Examples

```
data(pnadC_y2024q2)
attach(pnadC_y2024q2)
selectEq  <- participation ~ age + I(age^2) +
 male + white + yearsSchooling +
 hhold_head + hhold_spouse
outcomeEq <- ln_salary ~ age + I(age^2) +
 male + white + yearsSchooling +
 classWorker_employer + classWorker_selfEmployed
outcomeD  <- ~ age + I(age^2) +
 male + white + yearsSchooling +
 classWorker_employer + classWorker_selfEmployed
outcomeC  <- ~ male + yearsSchooling
fit_heckmanGE <- heckmanGE(selection  = selectEq,
                           outcome     = outcomeEq,
                           dispersion  = outcomeD,
                           correlation = outcomeC,
                           data = pnadC_y2024q2,
                           weights = weight,
                           cluster = ~PSU)
                           summary(fit_heckmanGE)
```

---

predict.heckmanGE   *predict.heckmanGE Predictions from the Generalized Heckman Model*

Description

Generates predictions from a fitted `heckmanGE` model. Predictions can be made on the scale of the linear predictors or on the scale of the response variable. The function can also return confidence intervals for the predictions if requested.

Usage

```
## S3 method for class 'heckmanGE'
predict(
  object,
  ...,
  part = c("selection", "outcome", "dispersion", "correlation"),
  newdata = NULL,
  type = c("link", "response"),
  cofint = FALSE,
  confidence_level = 0.95
)
```

Arguments

object   An object of class `heckmanGE`. This object should be a fitted model from which predictions will be made.

...	Argumentos adicionais passados para métodos específicos. Este argumento é mantido para compatibilidade com a função genérica `predict`.
part	A character vector specifying the model part for which to make predictions. Options include "selection", "outcome", "dispersion", and "correlation". The default is "outcome". If multiple parts are specified, only the "outcome" part will be used.
newdata	Optionally, a data frame containing new data for making predictions. If omitted, the function uses the fitted linear predictors from the model object.
type	The type of prediction required. The default is "link", which returns predictions on the scale of the linear predictors. If "response" is specified, predictions are returned on the scale of the response variable after applying the inverse link function.
cofint	A logical indicating whether to return confidence intervals for the predictions. Default is FALSE.
confidence_level	
	A numeric value specifying the confidence level for the confidence intervals if `cofint` is TRUE. Default is 0.95.

Details

- The function first checks the validity of the `part` and `type` arguments.
- If `newdata` is provided, the function ensures it matches the variables and structure of the original model frame.
- Predictions can be on the link scale or the response scale, depending on the `type` argument.
- Confidence intervals are calculated if `cofint` is TRUE, using the standard errors derived from the model.

Value

A vector or matrix of predictions from the `heckmanGE` object, depending on the value of `cofint`. If `cofint` is TRUE, the function returns a matrix with the mean predicted value, and the lower and upper bounds of the confidence interval.

---

print.heckmanGE	*print.heckmanGE Print of Generalized Heckman Model Results*

Description

Prints a summary of the results from a fitted `heckmanGE` model, including estimates for different model components, log-likelihood, AIC, BIC, and other relevant statistics.

Usage

```
## S3 method for class 'heckmanGE'
print(x, ...)
```

Arguments

x	An x of class `heckmanGE`. This x should be a fitted model whose results are to be printed.
...	Additional arguments passed to or from other methods.

Value

Prints the estimates and statistics of the Generalized Heckman model to the console.

`residuals.heckmanGE`	*residuals.heckmanGE Extract Residuals of the Generalized Heckman Model*

Description

Extracts residuals from a fitted `heckmanGE` model for a specified model component.

Usage

```
## S3 method for class 'heckmanGE'
residuals(object, part = c("selection", "outcome"), ...)
```

Arguments

object	An object of class `heckmanGE`. This object should be a fitted model from which residuals are to be extracted.
part	A character vector specifying which model component's residuals to return: either 'selection' or 'outcome'. Defaults to 'outcome'.
...	Additional arguments passed to or from other methods. These are not used in this method but must be included to match the generic method signature.

Value

A vector of residuals extracted from the specified part of the `heckmanGE` model.

---

sandwich.heckmanGE	*sandwich.heckmanGE Sandwich Estimator for Generalized Heckman Model*

Description

Computes the sandwich estimator of the variance-covariance matrix for the heckmanGE model.

Usage

```
sandwich.heckmanGE(x, bread. = bread.heckmanGE, meat. = meat.heckmanGE, ...)
```

Arguments

x	An object of class heckmanGE representing a fitted Generalized Heckman model.
bread.	A function to compute the "bread" part of the sandwich estimator. Defaults to bread.heckmanGE.
meat.	A function to compute the "meat" part of the sandwich estimator. Defaults to meat.heckmanGE.
...	Additional arguments passed to meat..

Value

A variance-covariance matrix for the heckmanGE model, computed using the sandwich estimator.

---

simulation	*Simulation dataset for the heckmanGE example*

Description

This dataset contains simulated data used to illustrate the functionality of the heckmanGE model. The data includes variables used in selection, outcome, dispersion, and correlation equations.

Usage

```
simulation
```

Format

A data frame with 10,000 observations on the following variables:

- y_o: Outcome variable from the simulated model (numeric)
- y_s: Selection indicator, 1 if selected, 0 otherwise (binary)
- prob_s: Probability of selection (numeric)

- x1: Simulated predictor from a normal distribution (numeric)
- x2: Simulated predictor from a Poisson distribution (numeric)
- x3: Simulated binary predictor (binary)
- x4: Simulated predictor from a normal distribution with mean 2 and sd 2 (numeric)
- x5: Simulated predictor from a Poisson distribution with lambda 1.5 (numeric)

Examples

```
data(simulation)
selectEq  <- y_s ~ x1 + x2 + x4
outcomeEq <- y_o ~ x1 + x2 + x3
outcomeD  <- ~ x1 + x5
outcomeC  <- ~ x3 + x4
fit_heckmanGE <- heckmanGE(selection   = selectEq,
                           outcome      = outcomeEq,
                           dispersion   = outcomeD,
                           correlation = outcomeC,
                           data = simulation)
summary(fit_heckmanGE)
```

---

step2   *step2: Two-Step Estimation Function*

Description

This function performs a two-step estimation process, commonly used in models that require correction for sample selection bias. It estimates parameters for selection, outcome, and dispersion equations.

Usage

```
step2(YS, XS, YO, XO, Msigma, Mrho, w)
```

Arguments

YS	A binary numeric vector indicating selection (1 if selected, 0 otherwise).
XS	A numeric matrix of covariates for the selection equation. Rows correspond to observations and columns to covariates.
YO	A numeric vector of observed outcomes for the selected sample (where YS == 1).
XO	A numeric matrix of covariates for the outcome equation. Rows correspond to selected observations.
Msigma	A numeric matrix of covariates for the dispersion equation.
Mrho	A numeric matrix of covariates for the correlation structure equation.
w	A numeric vector of weights to be used in the estimation process.

Details

This function implements a two-step estimation method for models with sample selection bias. The process begins by estimating the selection equation using a probit model to model the probability of selection. The Inverse Mills Ratio (IMR) is computed from the probit model and added as a covariate in the outcome and dispersion equations to correct for sample selection bias.

The outcome equation is estimated using weighted least squares (WLS), where the residuals are used to estimate the dispersion equation. Additionally, initial estimates for the correlation structure are computed based on the fitted values from the outcome equation.

Value

A list with the following elements:

selection	Estimated coefficients for the selection equation (probit model).
outcome	Estimated coefficients for the outcome equation (weighted least squares).
dispersion	Estimated coefficients for the dispersion equation (log of residual variance).
correlation	Initial guesses for the coefficients in the correlation structure.

See Also

[vcovHC](#) for computing robust standard errors.

---

summary.heckmanGE	*summary.heckmanGE Summary of Generalized Heckman Model*

Description

Provides a summary of the parameters and diagnostic information from a fitted Generalized Heckman model.

Usage

```
## S3 method for class 'heckmanGE'
summary(object, ...)
```

Arguments

object	An object of class heckmanGE representing a fitted Generalized Heckman model.
...	Additional arguments passed to other methods.

Value

Prints a detailed summary of the fitted Generalized Heckman model, including parameter estimates, standard errors, model fit statistics, and optimization details.

---

vcov.heckmanGE   *vcov.heckmanGE Variance-Covariance Matrix of the Generalized Heckman Model*

Description

Extracts the variance-covariance matrix of the coefficients for the Generalized Heckman model. The matrix can be for specific parts of the model or the complete matrix.

Usage

```
## S3 method for class 'heckmanGE'
vcov(
  object,
  part = c("selection", "outcome", "dispersion", "correlation"),
  ...
)
```

Arguments

object   An object of class heckmanGE representing a fitted Generalized Heckman model.

part   A character vector specifying the parts of the model to include in the variance-covariance matrix. Options are "selection", "outcome", "dispersion", and "correlation". By default, the function returns the complete variance-covariance matrix for all parts.

...   Additional arguments passed to or from other methods. These are not used in this method but must be included to match the generic method signature.

Value

A variance-covariance matrix of the coefficients for the specified parts of the Generalized Heckman model.

---

vcovCL.heckmanGE   *Variance-Covariance with Cluster Correction for Heckman Models*

Description

The vcovCL.heckmanGE function computes the variance-covariance matrix of a Heckman model, applying a cluster correction. This is useful for obtaining robust variance estimates, especially when there is within-group dependence.

Usage

```
vcovCL.heckmanGE(
  x,
  cluster = NULL,
  type = NULL,
  sandwich = TRUE,
  fix = FALSE,
  ...
)
```

Arguments

x	An object resulting from the estimation of a Heckman model using the `heckmanGE` method.
cluster	A vector or factor identifying clusters in the data. If NULL, assumes no clustering.
type	A character string specifying the type of cluster correction to be applied. It can be `"HC0"`, `"HC1"`, `"HC2"`, or `"HC3"`.
sandwich	A logical value. If TRUE, the function applies the sandwich estimator to the variance-covariance matrix.
fix	A logical value. If TRUE, corrects any negative eigenvalues in the variance-covariance matrix.
...	Additional arguments that can be passed to internal methods.

Details

This function is a specialized implementation for obtaining a robust variance-covariance matrix from Heckman models estimated with `heckmanGE`. It allows for cluster correction, which is particularly important in contexts where observations within groups may not be independent.

Value

A corrected variance-covariance matrix.

See Also

[meatCL.heckmanGE()](), [sandwich.heckmanGE()](), [bread.heckmanGE()]()

Index