

Package ‘medflex’

December 23, 2014

Title Flexible Mediation Analysis using Natural Effect Models

Version 0.4-6

Date 2014-12-23

Description medflex allows to run flexible mediation analyses using natural effect models as described in Lange, Vansteelandt and Bekaert (2012), Vansteelandt, Bekaert and Lange (2012) and Loeys, Moerkerke, De Smet, Buysse, Steen and Vansteelandt (2013).

Depends R (>= 3.1.2), multcomp (>= 1.3-6)

License GPL-2

LazyData true

Imports boot (>= 1.3-8)

Suggests arm (>= 1.7-05), car (>= 2.0-21), gam (>= 1.09.1), glmnet (>= 1.9-8), Matrix (>= 1.1-4), rpart (>= 4.1-8), SuperLearner (>= 2.0-15), VGAM (>= 0.9-5)

Author Johan Steen [aut, cre],
Tom Loeys [aut],
Beatrijs Moerkerke [aut],
Stijn Vansteelandt [aut],
Joris Meys [ctb] (technical support),
Theis Lange [ctb] (valuable suggestions)

Maintainer Johan Steen <johan.steen@gmail.com>

R topics documented:

expData	2
neImpute	3
neImpute.default	4
neImpute.formula	6
neLht	8
neLht-methods	10
neModel	11
neModel-methods	14
neWeight	16
neWeight.default	17
neWeight.formula	19
plot.neEffdecomp	21

plot.neModel 22

UPBdata 23

weights.expData 24

Index 25

expData	<i>Expanded dataset</i>
---------	-------------------------

Description

Expanded dataset including either ratio-of-mediator probability weights or imputed nested counterfactual outcomes.

Value

A data frame, resulting from applying [neWeight](#) or [neImpute](#) on an original dataset data. This data frame has `nRep * length(data)` rows, containing all original variables (except the original exposure variable) and two variables reflecting observed and hypothetical values of the exposure for each observation unit.

These auxiliary variables (x and x^*) are named after the exposure variable and carry integers as suffixes. Suffixes 0 and 1 are used for variables whose corresponding parameters in the final natural effect model index natural direct and indirect effects, respectively.

This object also stores some additional attributes, which are used as input for [neModel](#), such as

<code>model</code>	the fitted imputation model object
<code>data</code>	original dataset
<code>call</code>	the matched call
<code>terms</code>	the <code>neTerms</code> (internal class) object used
<code>weights</code>	ratio-of-mediator probability weights (only stored if object inherits from class <code>weightData</code>)

Note

If the weighting-based approach ([neWeight](#)) is applied, the original outcome values are copied for the nested counterfactual outcomes and the object stores an additional attribute, `"weights"`, containing a vector with ratio-of-mediator probability weights.

If the imputation-based approach ([neImpute](#)) is applied, the nested counterfactual outcomes are imputed by predictions from the imputation model.

In the former case, this object inherits from classes `c("data.frame", "expData", "impData")`, whereas in the latter case it inherits from classes `c("data.frame", "expData", "weightData")`.

See Also

[neImpute](#), [neWeight](#)

neImpute

Expand the dataset and impute nested counterfactual outcomes

Description

This function both expands the data along hypothetical exposure values and imputes nested counterfactual outcomes.

Usage

```
neImpute(object, ...)
```

Arguments

`object` an object used to select a method.
`...` additional arguments.

Details

Generic function that both expands the data along hypothetical exposure values (for each observation unit i) and imputes nested counterfactual outcomes in this expanded dataset in a single run. Imputed counterfactual outcomes

$$\hat{E}(Y_i | X_i = x, M_i, C_i)$$

are predictions from the imputation model that can be specified either externally as a fitted model object (`neImpute.default`) or internally (`neImpute.formula`).

Value

A data frame of class `c("data.frame", "expData", "impData")`. See [expData](#) for its structure.

References

Vansteelandt, S., Bekaert, M., & Lange, T. (2012). Imputation Strategies for the Estimation of Natural Direct and Indirect Effects. *Epidemiologic Methods*, **1**(1), Article 7.

Loeys, T., Moerkerke, B., De Smet, O., Buysse, A., Steen, J., & Vansteelandt, S. (2013). Flexible Mediation Analysis in the Presence of Nonlinear Relations: Beyond the Mediation Formula. *Multivariate Behavioral Research*, **48**(6), 871-894.

See Also

[neImpute.default](#), [neImpute.formula](#), [neModel](#), [expData](#)

neImpute.default

Expand the dataset and impute nested counterfactual outcomes

Description

This function both expands the data along hypothetical exposure values and imputes nested counterfactual outcomes.

Usage

```
## Default S3 method:
neImpute(object, formula, data, nMed = 1, nRep = 5,
  xSampling = c("quantiles", "random"), xFit, perCLim = c(0.05, 0.95), ...)
```

Arguments

object	fitted model object representing the imputation model.
formula	a formula object providing a symbolic description of the imputation model. Redundant if already specified in call for fitted model specified in object (see details).
data	data, as matrix or data frame, containing the exposure (and other relevant) variables. Redundant if already specified in call for fitted model specified in object (see details).
nMed	number of mediators.
nRep	number of replications or hypothetical values of the exposure to sample for each observation unit.
xSampling	character string indicating how to sample from the conditional exposure distribution. Possible values are "quantiles" or "random" (see details).
xFit	an optional fitted object (preferably glm) for the conditional exposure distribution (see details).
perCLim	a numerical vector of the form c(lower, upper) indicating the extreme percentiles to sample when using "quantiles" as sampling method to sample from the conditional exposure distribution (see details).
...	additional arguments.

Details

Imputed counterfactual outcomes are predictions from the imputation model that needs to be specified as a fitted object in the object argument.

If the model-fitting function used to fit the imputation model does not require specification of a formula or data argument (when using e.g. [SuperLearner](#)), these need to be specified explicitly in order to enable neImpute.default to extract pointers to variable types relevant for mediation analysis.

Whether a [formula](#) is specified externally (in the call for the fitted imputation model object which is specified in object) or internally (via the formula argument), it always needs to be of the form $Y \sim X + M1 + M2 + M3 + C1 + C2$, with the same outcome as in the final natural effect model and with predictor variables entered in the following prespecified order:

1. exposure X: The first predictor is coded as exposure or treatment.
2. mediator(s) M: The second predictor is coded as mediator. In case of multiple mediators ($nMed > 1$), then predictors 2: ($nMed + 1$) are coded as mediators.
3. baseline covariates C: All remaining predictor variables are automatically coded as baseline covariates.

If multiple mediators are specified ($nMed > 1$), the natural indirect effect parameter in the natural effect model captures the joint mediated effect. That is, the effect of the exposure on the outcome via these mediators considered jointly. The remaining effect of the exposure on the outcome (not mediated through the specified mediators) is then captured by the natural indirect effect parameter.

In contrast to imputation models with categorical exposures, additional arguments need to be specified if the exposure is continuous. All of these additional arguments are related to the sampling procedure for the exposure.

Whereas the number of replications `nRep` for categorical variables equals the number of levels for the exposure coded as a factor (i.e. the number of hypothetical exposure values), the number of desired replications needs to be specified explicitly for continuous exposures. Its default is 5.

If `xFit` is left unspecified, the hypothetical exposure levels are automatically sampled from a linear model for the exposure, conditional on a linear combination of all covariates. If one wishes to use another model for the exposure, this default model specification can be overruled by referring to a fitted model object in the `xFit` argument. Misspecification of this sampling model does not induce bias in the estimated coefficients and standard errors of the natural effect model.

The `xSampling` argument allows to specify how the hypothetical exposure levels should be sampled from the conditional exposure distribution (which is either entered explicitly using the `xFit` argument or fitted automatically as described in the previous paragraph). The "random" option randomly samples `nRep` draws from the exposure distribution, whereas the "quantiles" option (default) samples `nRep` quantiles at equal-sized probability intervals. Only the latter hence yields fixed exposure levels given `nRep` and `xFit`.

In order to guarantee that the entire support of the distribution is being sampled (which might be a concern if `nRep` is chosen to be small), the default lower and upper sampled quantiles are the 5th and 95th percentiles. The intermittent quantiles correspond to equal-sized probability intervals. So, for instance, if `nRep = 4`, then the sampled quantiles will correspond to probabilities 0.05, 0.35, 0.65 and 0.95. These default 'outer' quantiles can be changed by specifying the `perLim` argument accordingly. By specifying `perLim = NULL`, the standard quantiles will be sampled (e.g., 0.2, 0.4, 0.6 and 0.8 if `nRep = 4`).

Value

A data frame of class `c("data.frame", "expData", "impData")`. See [expData](#) for its structure.

See Also

[neImpute](#), [neImpute.formula](#), [neModel](#), [expData](#)

Examples

```
data(UPBdata)

## example using glm imputation model with binary exposure
fit.glm <- glm(UPB ~ attbin + negaff + gender + educ + age,
               family = binomial, data = UPBdata)
impData <- neImpute(fit.glm)
```

```

head(impData)

## example using glm imputation model with continuous exposure
fit.glm <- glm(UPB ~ att + negaff + gender + educ + age,
              family = binomial, data = UPBdata)
impData <- neImpute(fit.glm, nRep = 2)
head(impData)

## example using vglm (yielding identical results as with glm)
library(VGAM)
fit.vglm <- vglm(UPB ~ att + negaff + gender + educ + age,
                family = binomialff, data = UPBdata)
impData2 <- neImpute(fit.vglm, nRep = 2)
head(impData2)

## example using SuperLearner
library(Matrix)
library(SuperLearner)
SL.library <- c("SL.glm", "SL.glm.interaction", "SL.rpart",
               "SL.step", "SL.stepAIC", "SL.step.interaction",
               "SL.bayesglm", "SL.glmnet")
pred <- c("att", "negaff", "gender", "educ", "age")
fit.SL <- SuperLearner(Y = UPBdata$UPB, X = subset(UPBdata, select = pred),
                     SL.library = SL.library, family = binomial())
impSL <- neImpute(fit.SL,
                 formula = UPB ~ att + negaff + gender + educ + age,
                 data = UPBdata)
head(impSL)

```

neImpute.formula

Expand the dataset and impute nested counterfactual outcomes

Description

This function both expands the data along hypothetical exposure values and imputes nested counterfactual outcomes.

Usage

```

## S3 method for class 'formula'
neImpute(object, family, data, FUN = glm, nMed = 1,
        nRep = 5, xSampling = c("quantiles", "random"), xFit, perclim = c(0.05,
        0.95), ...)

```

Arguments

object	a formula object providing a symbolic description of the imputation model (see details).
family	a description of the error distribution and link function to be used in the model. Consult the help files of the model-fitting function specified in FUN for more details on appropriate argument specification.

data	data, as matrix or data frame, containing the exposure (and other relevant) variables. Redundant if already specified in call for fitted model specified in object (see details).
FUN	function used to fit model specified in formula.
nMed	number of mediators.
nRep	number of replications or hypothetical values of the exposure to sample for each observation unit.
xSampling	character string indicating how to sample from the conditional exposure distribution. Possible values are "quantiles" or "random" (see details).
xFit	an optional fitted object (preferably glm) for the conditional exposure distribution (see details).
percLim	a numerical vector of the form <code>c(lower, upper)</code> indicating the extreme percentiles to sample when using "quantiles" as sampling method to sample from the conditional exposure distribution (see details).
...	additional arguments (passed to FUN).

Details

Imputed counterfactual outcomes are predictions from the imputation model that is fitted internally by extracting information from the arguments `object`, `family`, `data`, `FUN` and ...

For imputation model specification via the `object` argument, use a [formula](#) of the form

$$Y \sim X + M1 + M2 + M3 + C1 + C2,$$

with the same outcome as in the final natural effect model and with predictor variables entered in the following prespecified order:

1. exposure X: The first predictor is coded as exposure or treatment.
2. mediator(s) M: The second predictor is coded as mediator. In case of multiple mediators ($nMed > 1$), then predictors 2: ($nMed + 1$) are coded as mediators.
3. baseline covariates C: All remaining predictor variables are automatically coded as baseline covariates.

If multiple mediators are specified ($nMed > 1$), the natural indirect effect parameter in the natural effect model captures the joint mediated effect. That is, the effect of the exposure on the outcome via these mediators considered jointly. The remaining effect of the exposure on the outcome (not mediated through the specified mediators) is then captured by the natural indirect effect parameter.

The type of imputation model can be defined by specifying an appropriate model-fitting function via the `FUN` argument (its default is [glm](#)). This method can only be used with model-fitting functions that require a `formula` argument (so not when using e.g. [SuperLearner](#)).

In contrast to imputation models with categorical exposures, additional arguments need to be specified if the exposure is continuous. All of these additional arguments are related to the sampling procedure for the exposure.

Whereas the number of replications `nRep` for categorical variables equals the number of levels for the exposure coded as a factor (i.e. the number of hypothetical exposure values), the number of desired replications needs to be specified explicitly for continuous exposures. Its default is 5.

If `xFit` is left unspecified, the hypothetical exposure levels are automatically sampled from a linear model for the exposure, conditional on a linear combination of all covariates. If one wishes to use another model for the exposure, this default model specification can be overruled by referring to a fitted model object in the `xFit` argument. Misspecification of this sampling model does not induce bias in the estimated coefficients and standard errors of the natural effect model.

The `xSampling` argument allows to specify how the hypothetical exposure levels should be sampled from the conditional exposure distribution (which is either entered explicitly using the `xFit` argument or fitted automatically as described in the previous paragraph). The "random" option randomly samples `nRep` draws from the exposure distribution, whereas the "quantiles" option (default) samples `nRep` quantiles at equal-sized probability intervals. Only the latter hence yields fixed exposure levels given `nRep` and `xFit`.

In order to guarantee that the entire support of the distribution is being sampled (which might be a concern if `nRep` is chosen to be small), the default lower and upper sampled quantiles are the 5th and 95th percentiles. The intermittent quantiles correspond to equal-sized probability intervals. So, for instance, if `nRep = 4`, then the sampled quantiles will correspond to probabilities 0.05, 0.35, 0.65 and 0.95. These default 'outer' quantiles can be changed by specifying the `perLim` argument accordingly. By specifying `perLim = NULL`, the standard quantiles will be sampled (e.g., 0.2, 0.4, 0.6 and 0.8 if `nRep = 4`).

Value

A data frame of class `c("data.frame", "expData", "impData")`. See [expData](#) for its structure.

See Also

[neImpute](#), [neImpute.default](#), [neModel](#), [expData](#)

Examples

```
data(UPBdata)

## example using glm imputation model with binary exposure
impData <- neImpute(UPB ~ attbin + negaff + gender + educ + age,
                    family = binomial, data = UPBdata)

head(impData)

## example using glm imputation model with continuous exposure
impData <- neImpute(UPB ~ att + negaff + gender + educ + age,
                    family = binomial, data = UPBdata, nRep = 2)

head(impData)

## example using vglm (yielding identical results as with glm)
library(VGAM)
impData2 <- neImpute(UPB ~ att + negaff + gender + educ + age,
                    family = binomialff, data = UPBdata,
                    nRep = 2, FUN = vglm)

head(impData2)
```

Description

`neLht` allows to calculate linear combinations of natural effect model parameter estimates.
`neEffdecomp` automatically extracts relevant causal parameter estimates from a natural effect model.

Usage

```
neEffdecomp(model, ...)

## S3 method for class 'neModel'
neEffdecomp(model, ...)

neLht(model, ...)

## S3 method for class 'neModelBoot'
neLht(model, ...)
```

Arguments

<code>model</code>	a fitted natural effect model object.
<code>...</code>	additional arguments (passed to glht).

Details

`neLht` is a wrapper of [glht](#) and offers the same functionality (see ‘Details’ section of [glht](#) for details on argument specification). It returns objects that inherit from the class “`neLht`” in order to make output of their corresponding methods (see [neLht-methods](#)) more compatible for natural effect models containing bootstrap variance-covariance matrices and standard errors.

`neEffdecomp` is a convenience function that automatically extracts causal parameter estimates from a natural effect model and derives natural effect components. That is, natural direct, natural indirect and total causal effect estimates are returned if no exposure-mediator interaction is modelled (i.e. two-way decomposition). If mediated interaction is allowed for in the natural effect model, there are two ways of decomposing the total effect into (natural) direct and indirect effects components: either as the sum of the pure direct and the total indirect effect or as the sum of the pure indirect and the total direct effect (i.e. three-way decomposition). In total, five causal effect estimates are returned in this case.

Value

An object of class `c("neLhtBoot", "neLht", "glht")` (see [glht](#)). `neEffdecomp` returns an object that additionally inherits from class “`neEffdecomp`”.

See [neLht-methods](#) for methods for `neLht` objects (and [glht-methods](#) for additional methods for `glht` objects).

Note

`neEffdecomp` is internally called by `plot.neModel` to create confidence interval plots for `neModel` objects.

See Also

[plot.neLht](#), [neLht-methods](#), [glht](#), [glht-methods](#), [neModel](#), [plot.neModel](#)

Examples

```
data(UPBdata)

impData <- neImpute(UPB ~ att * negaff + gender + educ + age,
  family = binomial, data = UPBdata)
```

```

neMod <- neModel(UPB ~ att0 * att1 + gender + educ + age,
  family = binomial, expData = impData)

lht <- neLht(neMod, linfct = c("att0 = 0", "att0 + att0:att1 = 0",
  "att1 = 0", "att1 + att0:att1 = 0",
  "att0 + att1 + att0:att1 = 0"))

summary(lht)

## or obtain directly via neEffdecomp
eff <- neEffdecomp(neMod)
summary(eff)

```

neLht-methods

Methods for linear hypotheses in natural effect models

Description

Obtain confidence intervals and statistical tests for linear hypotheses in natural effect models.

Usage

```

## S3 method for class 'neLhtBoot'
confint(object, parm, level = 0.95, type = "norm", ...)

## S3 method for class 'neLht'
summary(object, test = univariate(), ...)

```

Arguments

<code>object</code>	an object of class <code>neLht</code> .
<code>parm</code>	a specification of which parameters are to be given confidence intervals, either a vector of numbers or a vector of names. If missing, all parameters are considered.
<code>level</code>	the confidence level required.
<code>type</code>	the type of bootstrap intervals required. The default "norm" returns normal approximation bootstrap confidence intervals. Currently, only "norm", "basic" and "perc" are supported (see boot.ci).
<code>...</code>	additional arguments.
<code>test</code>	a function for computing p-values. The default <code>univariate()</code> does not apply a multiple testing correction. The function <code>adjusted()</code> allows to correct for multiple testing (see summary.glht and adjusted) and <code>Chisquare()</code> allows to test global linear hypotheses.

Details

`confint` yields bootstrap confidence intervals. These confidence intervals are internally called via the [boot.ci](#) function from the **boot** package (and not via the corresponding [confint.glht](#) function from the **multcomp** package!). The default confidence level specified in `level` (which corresponds to the `conf` argument in [boot.ci](#)) is 0.95 and the default type of bootstrap confidence interval, "norm", is based on the normal approximation (for more details see [boot.ci](#)).

A summary table with large sample tests, similar to that for [glht](#), can be obtained using `summary`.

In contrast to `summary.glht`, which by default returns p -values that are adjusted for multiple testing, the `summary` function returns unadjusted p -values. Adjusted p -values can also be obtained by specifying the `test` argument (see [adjusted](#) for more details).

Global Wald tests considering all linear hypotheses simultaneously (i.e. testing the global null hypothesis) can be requested by specifying `test = Chisqtest()`.

See [glht-methods](#) for additional methods for `glht` objects.

Note

Z-values in the summary table are simply calculated by dividing the parameter estimate by its corresponding bootstrap standard error. Corresponding p -values in the summary table are only indicative, since the null distribution for each statistic is assumed to be approximately standard normal. Therefore, where possible, it is generally recommended to focus mainly on bootstrap confidence intervals for inference, rather than the provided p -values.

See Also

[neLht](#), [plot.neLht](#), [glht](#), [glht-methods](#)

Examples

```
data(UPBdata)

impData <- neImpute(UPB ~ att * negaff + gender + educ + age,
  family = binomial, data = UPBdata)
neMod <- neModel(UPB ~ att0 * att1 + gender + educ + age,
  family = binomial, expData = impData)

lht <- neLht(neMod, linfct = c("att0 = 0", "att0 + att0:att1 = 0",
  "att1 = 0", "att1 + att0:att1 = 0",
  "att0 + att1 + att0:att1 = 0"))

## obtain bootstrap confidence intervals
confint(lht)
confint(lht, parm = c("att0", "att0 + att0:att1"))
confint(lht, parm = 1:2)
confint(lht, type = "perc", level = 0.90)

## summary table
summary(lht)

## summary table with omnibus Chisquare test
summary(lht, test = Chisqtest())
```

neModel

Fit a natural effect model

Description

`neModel` is used to fit a natural effect model on the expanded dataset.

Usage

```
neModel(formula, family = gaussian, expData, xFit, nBoot = 1000,
        parallel = c("no", "multicore", "snow"), ncpus = getOption("boot.ncpus",
        1L), progress = TRUE, ...)
```

Arguments

formula	a formula object providing a symbolic description of the natural effect model.
family	a description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See family for details of family functions.)
expData	the expanded dataset (of class " expData ").
xFit	fitted model object representing a model for the exposure (used for inverse treatment (exposure) probability weighting).
nBoot	number of bootstrap replicates (see R argument of boot).
parallel	The type of parallel operation to be used (if any). If missing, the default is taken from the option "boot.parallel" (and if that is not set, "no").
ncpus	integer: number of processes to be used in parallel operation: typically one would chose this to the number of available CPUs (see details).
progress	logical value indicating whether or not a progress bar should be displayed. Progress bars are automatically disabled for multicore processing.
...	additional arguments (passed to glm).

Details

This function is a wrapper for [glm](#), providing unbiased bootstrap standard errors for the parameter estimates.

The formula argument requires to be specified in function of the variables from the expanded dataset (specified in expData) whose corresponding parameters index the direct and indirect effect. Stratum-specific natural effects can be estimated by additionally modeling the relation between the outcome and baseline covariates. If the set of baseline covariates adjusted for in the formula argument is not sufficient to control for confounding (e.g. when fitting a population-average natural effect model), an adequate model for the exposure (conditioning on a sufficient set of baseline covariates) should be specified in the xFit argument. In this case, such a model for the exposure distribution is needed to weight by the reciprocal of the probability (density) of the exposure (i.e. inverse probability weighting) in order to adjust for confounding. Just as for ratio-of-mediator probability weighting (see paragraph below), this kind of weighting is done internally.

In contrast to [glm](#), the expData argument (rather than data argument) requires specification of a data frame that inherits from the class "[expData](#)", which contains additional information about e.g. the fitted working model, the variable types or terms of this working model and possibly ratio-of-mediator probability weights. The latter are automatically extracted from the [expData](#) object and weighting is done internally.

As the default [glm](#) standard errors fail to reflect the uncertainty inherent to the working model(s) (i.e. either a model for the mediator or an imputation model for the outcome and possibly a model for the exposure), bootstrap standard errors are calculated (using the [boot](#) function from the **boot** package). The bootstrap procedure entails refitting these working models on each bootstrap sample, reconstructing the expanded dataset and subsequently refitting the specified natural effect model on this dataset. In order to obtain stable standard errors, the number of bootstrap samples (specified

via the `nBoot` argument) should be chosen relatively high. Although the default is 1000, it is recommended to use at least 500 bootstrap samples.

To speed up the bootstrap procedure, parallel processing can be used by specifying the desired type of parallel operation via the `parallel` argument (for more details, see [boot](#)). The number of parallel processes (`ncpus`) is suggested to be specified explicitly (its default is 1, unless the global option `options("boot.cpus")` is specified). The function [detectCores](#) from the **parallel** package can be helpful at determining the number of available cores (although this may not always correspond to the number of *allowed* cores).

Value

An object of class "[neModel](#)" consisting of a list of 2 objects:

<code>neModelFit</code>	the fitted natural model object (of class " glm ") with downwardly biased standard errors
<code>bootRes</code>	the bootstrap results (of class " boot ")

See [neModel-methods](#) for methods for `neModel` objects.

Note

It is important to note that the original mediator(s) should not be specified in the `formula` argument, as the natural indirect effect in natural effect models should be captured solely by parameter(s) corresponding to the auxiliary hypothetical variable x^* in the expanded dataset (see [expData](#)).

References

- Lange, T., Vansteelandt, S., & Bekaert, M. (2012). A Simple Unified Approach for Estimating Natural Direct and Indirect Effects. *American Journal of Epidemiology*, **176**(3), 190-195.
- Vansteelandt, S., Bekaert, M., & Lange, T. (2012). Imputation Strategies for the Estimation of Natural Direct and Indirect Effects. *Epidemiologic Methods*, **1**(1), Article 7.
- Loeys, T., Moerkerke, B., De Smet, O., Buysse, A., Steen, J., & Vansteelandt, S. (2013). Flexible Mediation Analysis in the Presence of Nonlinear Relations: Beyond the Mediation Formula. *Multivariate Behavioral Research*, **48**(6), 871-894.

See Also

[neModel-methods](#), [plot.neModel](#), [neImpute](#), [neWeight](#), [neLht](#), [neEffdecomp](#)

Examples

```
data(UPBdata)

## weighting-based approach
weightData <- neWeight(negaffect ~ att + gender + educ + age,
                      data = UPBdata)

# stratum-specific natural effects
weightFit1 <- neModel(UPB ~ att0 * att1 + gender + educ + age,
                    family = binomial, expData = weightData)
summary(weightFit1)

# population-average natural effects
expFit <- glm(att ~ gender + educ + age, data = UPBdata)
```

```

weightFit2 <- neModel(UPB ~ att0 * att1, family = binomial,
                     expData = weightData, xFit = expFit)
summary(weightFit2)

## imputation-based approach
impData <- neImpute(UPB ~ att * negaff + gender + educ + age,
                  family = binomial, data = UPBdata)

# stratum-specific natural effects
impFit1 <- neModel(UPB ~ att0 * att1 + gender + educ + age,
                  family = binomial, expData = impData)
summary(impFit1)

# population-average natural effects
impFit2 <- neModel(UPB ~ att0 * att1, family = binomial,
                  expData = impData, xFit = expFit)
summary(impFit2)

```

neModel-methods

Methods for natural effect models

Description

Extractor functions, confidence intervals and statistical tests for natural effect models.

Usage

```

## S3 method for class 'neModel'
coef(object, ...)

## S3 method for class 'neModelBoot'
confint(object, parm, level = 0.95, type = "norm",
        ...)

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

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

## S3 method for class 'neModel'
weights(object, ...)

```

Arguments

object	a fitted natural effect model object.
...	additional arguments (see boot.ci for <code>confint</code> or summary.glm for <code>summary</code>).
parm	a specification of which parameters are to be given confidence intervals, either a vector of numbers or a vector of names. If missing, all parameters are considered.

level	the confidence level required.
type	the type of bootstrap intervals required. The default "norm" returns normal approximation bootstrap confidence intervals. Currently, only "norm", "basic" and "perc" are supported (see boot.ci).

Details

confint yields bootstrap confidence intervals. These confidence intervals are internally called via the [boot.ci](#) function from the **boot** package. The default confidence level specified in level (which corresponds to the conf argument in [boot.ci](#)) is 0.95 and the default type of bootstrap confidence interval, "norm", is based on the normal approximation (for more details see [boot.ci](#)).

A summary table with large sample tests, similar to that for [glm](#) output, can be obtained using summary.

vcov returns the variance covariance matrix calculated from the bootstrap samples stored in object\$bootRes (see [neModel](#)).

weights returns a vector containing the regression weights used to fit the natural effect model. These weights can be based on

1. ratio-of-mediator probability (density) weights (only if the weighting-based approach is used)
2. inverse probability of treatment (exposure) weights (only if xFit was specified in [neModel](#))

Note

Z-values in the summary table are simply calculated by dividing the parameter estimate by its corresponding bootstrap standard error. Corresponding *p*-values in the summary table are only indicative, since the null distribution for each statistic is assumed to be approximately standard normal. Therefore, where possible, it is generally recommend to focus mainly on bootstrap confidence intervals for inference, rather than the provided *p*-values.

See Also

[neModel](#), [plot.neModel](#), [weights](#)

Examples

```
data(UPBdata)

weightData <- neWeight(negaffect ~ att + educ + gender + age,
                      data = UPBdata)
neMod <- neModel(UPB ~ att0 * att1 + educ + gender + age,
                family = binomial, expData = weightData)

## extract coefficients
coef(neMod)

## extract (bootstrap) variance-covariance matrix
vcov(neMod)

## extract regression weights
w <- weights(neMod)
head(w)

## obtain bootstrap confidence intervals
confint(neMod)
```

```

confint(neMod, parm = c("att0"))
confint(neMod, type = "perc", level = 0.90)

## summary table
summary(neMod)

```

neWeight	<i>Expand the dataset and calculate ratio-of-mediator probability weights</i>
----------	---

Description

This function both expands the data along hypothetical exposure values and calculates ratio-of-mediator probability weights.

Usage

```
neWeight(object, ...)
```

Arguments

object	an object used to select a method.
...	additional arguments.

Details

Generic function that both expands the data along hypothetical exposure values and calculates ratio-of-mediator probability weights

$$\frac{\hat{P}(M_i|X_i = x^*, C_i)}{\hat{P}(M_i|X_i = x, C_i)}$$

for each observation unit i in this expanded dataset in a single run. These weights are ratios of probabilities or probability densities from the mediator model distribution, which can be specified either externally as a fitted model object ([neWeight.default](#)) or internally ([neWeight.formula](#)).

Value

A data frame of class `c("data.frame", "expData", "weightData")`. See [expData](#) for its structure.

References

Lange, T., Vansteelandt, S., & Bekaert, M. (2012). A Simple Unified Approach for Estimating Natural Direct and Indirect Effects. *American Journal of Epidemiology*, **176**(3), 190-195.

See Also

[neWeight.default](#), [neWeight.formula](#), [expData](#)

neWeight.default	<i>Expand the dataset and calculate ratio-of-mediator probability weights</i>
------------------	---

Description

This function both expands the data along hypothetical exposure values and calculates ratio-of-mediator probability weights.

Usage

```
## Default S3 method:
neWeight(object, formula, data, nRep = 5,
  xSampling = c("quantiles", "random"), xFit, perCLim = c(0.05, 0.95), ...)
```

Arguments

object	fitted model object representing the mediator model.
formula	a formula object providing a symbolic description of the mediator model. Redundant if already specified in call for fitted model specified in object (see details).
data	data, as matrix or data frame, containing the exposure (and other relevant) variables. Redundant if already specified in call for fitted model specified in object (see details).
nRep	number of replications or hypothetical values of the exposure to sample for each observation unit.
xSampling	character string indicating how to sample from the conditional exposure distribution. Possible values are "quantiles" or "random" (see details).
xFit	an optional fitted object (preferably glm) for the conditional exposure distribution (see details).
perCLim	a numerical vector of the form c(lower, upper) indicating the extreme percentiles to sample when using "quantiles" as sampling method to sample from the conditional exposure distribution (see details).
...	additional arguments.

Details

The calculated weights are ratios of fitted probabilities or probability densities from the distribution of the mediator model. This model needs to be specified as a fitted object in the object argument.

If the model-fitting function used to fit the mediator model does not require specification of a formula or data argument, these need to be specified explicitly in order to enable neWeight.default to extract pointers to variable types relevant for mediation analysis.

Whether a [formula](#) is specified externally (in the call for the fitted mediator model object which is specified in object) or internally (via the formula argument), it always needs to be of the form $M \sim X + C1 + C2$, with predictor variables entered in the following prespecified order:

1. exposure X: The first predictor is coded as exposure or treatment.
2. baseline covariates C: All remaining predictor variables are automatically coded as baseline covariates.

In contrast to imputation models with categorical exposures, additional arguments need to be specified if the exposure is continuous. All of these additional arguments are related to the sampling procedure for the exposure.

Whereas the number of replications `nRep` for categorical variables equals the number of levels for the exposure coded as a factor (i.e. the number of hypothetical exposure values), the number of desired replications needs to be specified explicitly for continuous exposures. Its default is 5.

If `xFit` is left unspecified, the hypothetical exposure levels are automatically sampled from a linear model for the exposure, conditional on a linear combination of all covariates. If one wishes to use another model for the exposure, this default model specification can be overruled by referring to a fitted model object in the `xFit` argument. Misspecification of this sampling model does not induce bias in the estimated coefficients and standard errors of the natural effect model.

The `xSampling` argument allows to specify how the hypothetical exposure levels should be sampled from the conditional exposure distribution (which is either entered explicitly using the `xFit` argument or fitted automatically as described in the previous paragraph). The "random" option randomly samples `nRep` draws from the exposure distribution, whereas the "quantiles" option (default) samples `nRep` quantiles at equal-sized probability intervals. Only the latter hence yields fixed exposure levels given `nRep` and `xFit`.

In order to guarantee that the entire support of the distribution is being sampled (which might be a concern if `nRep` is chosen to be small), the default lower and upper sampled quantiles are the 5th and 95th percentiles. The intermittent quantiles correspond to equal-sized probability intervals. So, for instance, if `nRep` = 4, then the sampled quantiles will correspond to probabilities 0.05, 0.35, 0.65 and 0.95. These default 'outer' quantiles can be changed by specifying the `perCLim` argument accordingly. By specifying `perCLim` = NULL, the standard quantiles will be sampled (e.g., 0.2, 0.4, 0.6 and 0.8 if `nRep` = 4).

Value

A data frame of class `c("data.frame", "expData", "weightData")`. See [expData](#) for its structure.

See Also

[neWeight](#), [neWeight.formula](#), [expData](#)

Examples

```
data(UPBdata)

## example using glm
fit.glm <- glm(negaff ~ att + gender + educ + age, data = UPBdata)
weightData <- neWeight(fit.glm, nRep = 2)

## example using vglm (yielding identical results as with glm)
library(VGAM)
fit.vglm <- vglm(negaff ~ att + gender + educ + age,
                 family = gaussianff, data = UPBdata)
weightData2 <- neWeight(fit.vglm, nRep = 2)
```

neWeight.formula	<i>Expand the dataset and calculate ratio-of-mediator probability weights</i>
------------------	---

Description

This function both expands the data along hypothetical exposure values and calculates ratio-of-mediator probability weights.

Usage

```
## S3 method for class 'formula'
neWeight(object, family, data, FUN = glm, nRep = 5,
  xSampling = c("quantiles", "random"), xFit, perLim = c(0.05, 0.95), ...)
```

Arguments

object	a formula object providing a symbolic description of the mediator model (see details).
family	a description of the error distribution and link function to be used in the model. Consult the help files of the model-fitting function specified in FUN for more details on appropriate argument specification.
data	data, as matrix or data frame, containing the exposure (and other relevant) variables. Redundant if already specified in call for fitted model specified in object (see details).
FUN	function used to fit model specified in formula.
nRep	number of replications or hypothetical values of the exposure to sample for each observation unit.
xSampling	character string indicating how to sample from the conditional exposure distribution. Possible values are "quantiles" or "random" (see details).
xFit	an optional fitted object (preferably glm) for the conditional exposure distribution (see details).
perLim	a numerical vector of the form c(lower, upper) indicating the extreme percentiles to sample when using "quantiles" as sampling method to sample from the conditional exposure distribution (see details).
...	additional arguments (passed to FUN).

Details

The calculated weights are ratios of fitted probabilities or probability densities from the distribution of the mediator model. This model is fitted internally by extracting information from the arguments object, family, data, FUN and

For mediation model specification via the object argument, use a [formula](#) of the form $M \sim X + C1 + C2$, with predictor variables entered in the following prespecified order:

1. exposure X: The first predictor is coded as exposure or treatment.
2. baseline covariates C: All remaining predictor variables are automatically coded as baseline covariates.

The type of mediator model can be defined by specifying an appropriate model-fitting function via the FUN argument (its default is [glm](#)). This method can only be used with model-fitting functions that require a formula argument.

In contrast to imputation models with categorical exposures, additional arguments need to be specified if the exposure is continuous. All of these additional arguments are related to the sampling procedure for the exposure.

Whereas the number of replications nRep for categorical variables equals the number of levels for the exposure coded as a factor (i.e. the number of hypothetical exposure values), the number of desired replications needs to be specified explicitly for continuous exposures. Its default is 5.

If xFit is left unspecified, the hypothetical exposure levels are automatically sampled from a linear model for the exposure, conditional on a linear combination of all covariates. If one wishes to use another model for the exposure, this default model specification can be overruled by referring to a fitted model object in the xFit argument. Misspecification of this sampling model does not induce bias in the estimated coefficients and standard errors of the natural effect model.

The xSampling argument allows to specify how the hypothetical exposure levels should be sampled from the conditional exposure distribution (which is either entered explicitly using the xFit argument or fitted automatically as described in the previous paragraph). The "random" option randomly samples nRep draws from the exposure distribution, whereas the "quantiles" option (default) samples nRep quantiles at equal-sized probability intervals. Only the latter hence yields fixed exposure levels given nRep and xFit.

In order to guarantee that the entire support of the distribution is being sampled (which might be a concern if nRep is chosen to be small), the default lower and upper sampled quantiles are the 5th and 95th percentiles. The intermittent quantiles correspond to equal-sized probability intervals. So, for instance, if nRep = 4, then the sampled quantiles will correspond to probabilities 0.05, 0.35, 0.65 and 0.95. These default 'outer' quantiles can be changed by specifying the perLim argument accordingly. By specifying perLim = NULL, the standard quantiles will be sampled (e.g., 0.2, 0.4, 0.6 and 0.8 if nRep = 4).

Value

A data frame of class `c("data.frame", "expData", "weightData")`. See [expData](#) for its structure.

See Also

[neWeight.default](#), [expData](#)

Examples

```
data(UPBdata)

## example using glm
weightData <- neWeight(negaffect ~ att + gender + educ + age,
                      data = UPBdata, nRep = 2)

## example using vglm (yielding identical results as with glm)
library(VGAM)
weightData2 <- neWeight(negaffect ~ att + gender + educ + age,
                      family = gaussianff, data = UPBdata, nRep = 2, FUN = vglm)
```

plot.neEffdecomp	<i>Confidence interval plots for linear hypotheses in natural effect models</i>
------------------	---

Description

Confidence interval plots for linear hypotheses in natural effect models.

Usage

```
## S3 method for class 'neEffdecomp'
plot(x, level = 0.95, ci.type = "norm",
     transf = identity, ylabels, yticks.at, ...)

## S3 method for class 'neLht'
plot(x, level = 0.95, ci.type = "norm", transf = identity,
     ylabels, yticks.at, ...)
```

Arguments

x	an object of class neLht.
level	the confidence level required.
ci.type	the type of bootstrap intervals required. The default "norm" returns normal approximation bootstrap confidence intervals. Currently, only "norm", "basic" and "perc" are supported (see boot.ci).
transf	transformation function to be applied internally on the (linear hypothesis) estimates and their confidence intervals (e.g. exp for logit or Poisson regression). The default is identity (i.e. no transformation).
ylabels	character vector containing the labels for the (linear hypothesis) estimates to be plotted on the y-axis.
yticks.at	numeric vector containing the y-coordinates (from 0 to 1) to draw the tick marks for the different estimates and their corresponding confidence intervals.
...	additional arguments.

Details

This function is an adapted version of [plot.glht](#) from the **multcomp** package and yields confidence interval plots for each of the linear hypothesis parameters.

See Also

[neModel](#), [neEffdecomp](#), [plot.neLht](#), [plot.neEffdecomp](#)

Examples

```
data(UPBdata)

impData <- neImpute(UPB ~ att * negaff + gender + educ + age,
                   family = binomial, data = UPBdata)
neMod <- neModel(UPB ~ att0 * att1 + gender + educ + age,
                family = binomial, expData = impData)
```

```

lht <- neLht(neMod, linfct = c("att0 = 0", "att0 + att0:att1 = 0",
                              "att1 = 0", "att1 + att0:att1 = 0",
                              "att0 + att1 + att0:att1 = 0"))

## all pairs return identical output
plot(confint(lht), transf = exp)
plot(lht, transf = exp)

plot(neEffdecomp(neMod), transf = exp)
plot(neMod, transf = exp)

```

plot.neModel	<i>Confidence interval plots for natural effect estimates</i>
--------------	---

Description

Obtain effect decomposition confidence interval plots for natural effect models.

Usage

```

## S3 method for class 'neModel'
plot(x, level = 0.95, ci.type = "norm",
     transf = identity, ylabels, yticks.at, ...)

```

Arguments

x	a fitted natural effect model object.
level	the confidence level required.
ci.type	the type of bootstrap intervals required. The default "norm" returns normal approximation bootstrap confidence intervals. Currently, only "norm", "basic" and "perc" are supported (see boot.ci).
transf	transformation function to be applied internally on the (linear hypothesis) estimates and their confidence intervals (e.g. exp for logit or Poisson regression). The default is identity (i.e. no transformation).
ylabels	character vector containing the labels for the (linear hypothesis) estimates to be plotted on the y-axis.
yticks.at	numeric vector containing the y-coordinates (from 0 to 1) to draw the tick marks for the different estimates and their corresponding confidence intervals.
...	additional arguments.

Details

This function yields confidence interval plots for the natural effect parameters of interest. These causal parameter estimates are first internally extracted from the neModel object by applying [neEffdecomp](#).

Examples

```
data(UPBdata)

impData <- neImpute(UPB ~ att * negaff + educ + gender + age,
                   family = binomial, data = UPBdata)
neMod <- neModel(UPB ~ att0 * att1 + educ + gender + age,
                family = binomial, expData = impData)

plot(neMod)
plot(neMod, transf = exp,
     ylabels = c("PDE", "TDE", "PIE", "TIE", "TE"))
```

UPBdata	<i>UPB data</i>
---------	-----------------

Description

Data from a survey study that was part of the Interdisciplinary Project for the Optimization of Separation trajectories (IPOS). This large-scale project involved the recruitment of individuals who divorced between March 2008 and March 2009 in four major courts in Flanders. It aimed to improve the quality of life in families during and after the divorce by translating research findings into practical guidelines for separation specialists and by promoting evidence-based policy. This dataset involves a subsample of 385 individuals, namely those who responded to a battery of questionnaires related to romantic relationship and breakup characteristics (De Smet, 2012).

Format

A data frame with 385 rows and 9 variables:

att self-reported anxious attachment level (standardized)

attbin binary version of self-reported anxious attachment level: 1 = higher than sample mean, 0 = lower than sample mean

negaff level of self-reported experienced negative affectivity (standardized)

negaffcat multicategorical version of level of self-reported experienced negative affectivity: L = low, M = intermediate, H = high

initiator initiator of the divorce

gender gender: F = female, M = male

educ education level: either H = high (at least a bachelor's degree), M = intermediate (having finished secondary school) or L = low (otherwise)

age age (in years)

UPB binary variable indicating whether the individual reported having displayed unwanted pursuit behavior(s) towards the ex-partner

Source

Ghent University and Catholic University of Louvain (2010). *Interdisciplinary Project for the Optimisation of Separation trajectories - divorce and separation in Flanders*.

<http://www.scheidingsonderzoek.ugent.be/index-eng.html>

References

- De Smet, O., Loeys, T., & Buysse, A. (2012). Post-Breakup Unwanted Pursuit: A Refined Analysis of the Role of Romantic Relationship Characteristics. *Journal of Family Violence*, **27**(5), 437-452.
- Loeys, T., Moerkerke, B., De Smet, O., Buysse, A., Steen, J., & Vansteelandt, S. (2013). Flexible Mediation Analysis in the Presence of Nonlinear Relations: Beyond the Mediation Formula. *Multivariate Behavioral Research*, **48**(6), 871-894.

weights.expData	<i>Extract regression weights from the expanded dataset</i>
-----------------	---

Description

This function extracts the regression weights (to be used in the natural effect model) for each observation of an expanded dataset.

Usage

```
## S3 method for class 'expData'
weights(object, ...)
```

Arguments

object	an expanded dataset (of class " expData ").
...	additional arguments.

Value

a vector of length `nrow(object)`, containing the regression weights for the expanded dataset specified in object.

See Also

[expData](#), [neWeight](#), [weights](#)

Examples

```
data(UPBdata)

weightData <- neWeight(negaff ~ att + gender + educ + age,
                      data = UPBdata, nRep = 2)

head(weights(weightData))
```


Index

adjusted, [10](#), [11](#)

boot, [12](#), [13](#)
boot.ci, [10](#), [14](#), [15](#), [21](#), [22](#)

coef.neModel (neModel-methods), [14](#)
confint.glht, [10](#)
confint.neLhtBoot (neLht-methods), [10](#)
confint.neModelBoot (neModel-methods),
[14](#)

detectCores, [13](#)

expData, [2](#), [3](#), [5](#), [8](#), [12](#), [13](#), [16](#), [18](#), [20](#), [24](#)

family, [12](#)
formula, [4](#), [6](#), [7](#), [12](#), [17](#), [19](#)

glht, [9](#), [11](#)
glm, [7](#), [12](#), [13](#), [15](#), [20](#)

neEffdecomp, [13](#), [21](#), [22](#)
neEffdecomp (neLht), [8](#)
neImpute, [2](#), [3](#), [5](#), [8](#), [13](#)
neImpute.default, [3](#), [4](#), [8](#)
neImpute.formula, [3](#), [5](#), [6](#)
neLht, [8](#), [11](#), [13](#)
neLht-methods, [10](#)
neModel, [2](#), [3](#), [5](#), [8](#), [9](#), [11](#), [13](#), [15](#), [21](#)
neModel-methods, [14](#)
neWeight, [2](#), [13](#), [16](#), [18](#), [24](#)
neWeight.default, [16](#), [17](#), [20](#)
neWeight.formula, [16](#), [18](#), [19](#)

plot.glht, [21](#)
plot.neEffdecomp, [21](#), [21](#)
plot.neLht, [9](#), [11](#), [21](#)
plot.neLht (plot.neEffdecomp), [21](#)
plot.neModel, [9](#), [13](#), [15](#), [22](#)

summary.glht, [10](#), [11](#)
summary.glm, [14](#)
summary.neLht (neLht-methods), [10](#)
summary.neModel (neModel-methods), [14](#)
SuperLearner, [4](#), [7](#)

UPBdata, [23](#)

vcov.neModelBoot (neModel-methods), [14](#)

weights, [15](#), [24](#)
weights.expData, [24](#)
weights.neModel (neModel-methods), [14](#)