Package 'collin'

February 15, 2021

Type Package

Title Effects of Collinearity in Distributed Lag and Other Models

Version 0.0.1

Description Visual tool to assessing whether the results of a study could be driven by collinearity. The methods are described in Basagana X, Barrera-Gomez J (2021) ``Visualizing the effects of collinearity in distributed lag models". International Journal of Epidemiology (under review).

Depends R (>= 3.6.0)

License GPL-3

Encoding UTF-8

LazyData true

Imports dlnm, graphics, grDevices, MASS, mgcv, nlme, stats, utils, VGAM

Suggests knitr, rmarkdown, splines, xtable

RoxygenNote 6.1.1

VignetteBuilder knitr

NeedsCompilation no

Author Jose Barrera-Gomez [aut, cre], Xavier Basagana [aut]

Maintainer Jose Barrera-Gomez <jose.barrera@isglobal.org>

Repository CRAN

Date/Publication 2021-02-15 16:30:02 UTC

R topics documented:

`collin` *collin: Visualizing the effects of collinearity in distributed lag models and other linear models*

Description

Visual tool to assessing whether the results of a study could be driven by collinearity.

Procedure

The tool consists of a two-step procedure.

In the first step, the user provides the fitted model to be analyzed and a hypothetical effect pattern of the predictor of interest on the outcome. Then, simulations are performed under such a hypothetical effects and the given model.

In the second step, a specific `plot` method is used to visualize the results of the simulations to assessing whether these results are consistent with the results using the real data, so that unexpected results could be driven by collinearity.

Authors

Jose Barrera-Gomez and Xavier Basagana Maintainer: Jose Barrera-Gomez <jose.barrera@isglobal.org>

References

The methodology used in the package is described in

Basagana X, Barrera-Gomez J. Visualizing the effects of collinearity in distributed lag models. *International Journal of Epidemiology*. (under review)

`collindlnm` *Simulates effects from a distributed lag model pattern.*

Description

Simulates results from a distributed lag model under an hypothetical effect pattern provided by the user, which can be linear or non-linear. The output is the passed to the `plot` method to visualize consequences of collinearity.

Usage

```
collindlnm(model, x, cb, at = 1, cen = 0, effect, type = c("coef",
  "risk"), shape = c("linear", "nonlinear"), nsim = 100,
  verbose = TRUE, seed = NULL)
```

Arguments

model	a model that includes a crossbasis. Currently, models allowed are those of class "glm" or "lme", if shape = "linear"; or "glm", if shape = "nonlinear".
x	if shape = "linear", a matrix that includes the values of the predictor under study, in the first column, and the lagged values, up to the maximum lag considered, as the subsequent columns. If shape = "nonlinear", a numeric vector or 1-column matrix including the original (i.e. lag 0) values of the predictor.
cb	an object of class "crossbasis". The crossbasis included in model.
at	the increase(s) in the predictor under study to be considered to report the effects of the variable. If shape = "linear", at must be a single number. If shape = "nonlinear", at must be a numeric vector with at least two different values.
cen	a number. Reference value of the predictor under study, used to calculate effects. If shape = "linear", the value of cen is irrelevant (and it is internally set to 0).
effect	a vector or a matrix, depending on shape, including the hypothetical effect of the predictor under analysis. If shape = "linear", a vector including the linear effect at each lag (including lag 0). If shape = "nonlinear", a matrix including the effect at each lag (including lag 0) (columns) and for each value in at (rows).
type	a character. If type = "coef" (default), effect is supposed to be in the linear predictor scale (i.e. it is considered as regression coefficients in model). If type = "risk", effect is supposed to be in terms of relative risks (i.e. exp(coef), as ORs or RRs in logistic or Poisson families, respectively). If model is of class "lme", then it must be type = "coef" (default).
shape	the shape of the relationship between the linear predictor of the model and the outcome. Default is "linear".
nsim	number of simulations. Default is 100.
verbose	a logical value indicating output status messages. Default is TRUE.
seed	a number. Seed for reproducibility of results. Default is NULL (no seed).

Value

A list including the results of the simulations to be passed to the plot method.

See Also

crossbasis, glm, lme.

Examples

```
# For detailed examples:
browseVignettes("collin")
```

lagpad *Function to get a vector lagged.*

Description

Function to get a vector lagged.

Usage

```
lagpad(x, k)
```

Arguments

x	numeric vector to be lagged.
k	number of lags to be applied.

Value

A lagged vector.

Examples

```
s <- 1:5
s1 <- lagpad(s, k = 1)
s2 <- lagpad(s, k = 2)
data.frame(s, s1, s2)
```

mempm25 *Working memory test scores and air pollution.*

Description

Simulated data for repeated measures of working memory test scores and residential $PM_{2.5}$ levels in 2,221 children.

Usage

```
mempm25
```

Format

A data frame with 8,884 rows and 17 variables:

id individual identifier

session test session identifier

school chool identifier

sex sex of individual

agecen age of individual, centered at the mean, in years

educ maternal education

resses residential neighborhood socioeconomic status indicator

pm25y0 estimated average concentration of $PM_{2.5}$ ($\mu g/m^3$) at residence address during pregnancy

pm25y1 estimated average concentration of $PM_{2.5}$ ($\mu g/m^3$) at residence address during year 1

pm25y2 estimated average concentration of $PM_{2.5}$ ($\mu g/m^3$) at residence address during year 2

pm25y3 estimated average concentration of $PM_{2.5}$ ($\mu g/m^3$) at residence address during year 3

pm25y4 estimated average concentration of $PM_{2.5}$ ($\mu g/m^3$) at residence address during year 4

pm25y5 estimated average concentration of $PM_{2.5}$ ($\mu g/m^3$) at residence address during year 5

pm25y6 estimated average concentration of $PM_{2.5}$ ($\mu g/m^3$) at residence address during year 6

pm25y7 estimated average concentration of $PM_{2.5}$ ($\mu g/m^3$) at residence address during year 7

wei weight of the observation

wmemo working memory test score

Details

This is a synthetic dataset generated with the `synthpop` package, based on a cohort study of children examining the association between $PM_{2.5}$ concentrations during the prenatal period and the first seven postnatal years on cognitive tests taken at age 8 (see 'References').

Source

See 'References'.

References

Rivas I *et al*. Association between Early Life Exposure to Air Pollution and Working Memory and Attention. *Environmental Health Perspectives*. 2019;127(5):57002. doi: 10.1289/EHP3169.

Examples

```
# correlation of air pollution levels between two different periods:
pm25yearly <- mempm25[, grep("pm25", names(mempm25))]
print(cor(pm25yearly, use = "complete.obs"), digits = 2)
```

plot.collindlnmlinear *Visualizes effects of collinearity in distributed lag model under an hypothetical linear effect pattern.*

Description

Visualize the results from a distributed lag model under an hypothetical linear effect pattern provided by the user, generated using the function collindlnm.

Usage

```
## S3 method for class 'collindlnmlinear'
plot(x, lags = NULL, ...)
```

Arguments

x	an object of class "collindlnmlinear", which is generated by the function collindlnm with shape = "linear".
lags	a number or a numeric vector indicating at what lags the results will be visualized. Default (NULL) shows all lags.
...	other parameters to be passed through to plot function.

Value

A plot showing a comparison between the results under the fitted model and the results under the hypothetical true effect.

See Also

collindlnm, plot.collindlnmnonlinear.

Examples

```
# For detailed examples:
browseVignettes("collin")
```

plot.collindlnmnonlinear *Visualize effects of collinearity in distributed lag model under an hypothetical non-linear effect pattern.*

Description

Visualize the results from a distributed lag model under an hypothetical non-linear effect pattern provided by the user, generated using the function collindlnm. The number of plots shown is equal to the number of values passed by at in the function collindlnm. The way in which these plots are displayed is controlled by the user through the argument show.

Usage

```
## S3 method for class 'collindlnmnonlinear'
plot(x, lags = NULL, show = c("manual",
  "auto", "sequence"), addlegend = TRUE, varlegend = NULL, ...)
```

Arguments

x	an object of class "collindlnmnonlinear", which is generated by the function collindlnm with shape = "nonlinear".
lags	a number or a numeric vector indicating at what lags the results will be visualized. Default (NULL) shows all lags.
show	character indicating how the multiple plots will be shown. If show = "manual", then it is expected that the user have previously set the graphical parameters to arrange them (i.e. setting mfrow) with par. If show = "auto", then the arrangement of the plots (i.e. the value of mfrow) is automatically set. If show = "sequence", then the plots are sequentially overlaid.
addlegend	logical indicating whether a legend indicating at what value (of at passed in collindlnm) the results correspond to.
varlegend	character indicating the label for the explored variable to be shown in the legend.
...	other parameters to be passed through to plot function.

Value

A plot showing a comparison between results under the fitted model and the results under the hypothetical true effect, for each of the different values of the variable of interest where effects were explored.

See Also

collindlnm, plot.collindlnmlinear.

---

rhospno2 *Respiratory hospital admissions and air pollution.*

Description

Simulated data for daily measures of hospital admissions for respiratory causes count and ambient air NO_2 concentrations during 10 years.

Usage

```
rhospno2
```

Format

A data frame with 3,652 rows and 7 variables:

date date of the observation

t numerical indicator of date

year year indicator

dow day of week

temp ambient temperature in Celsius degrees

no2 NO_2 concentration in $\mu g/m^3$

hresp number of hospital admissions for respiratory causes

Details

This is a synthetic dataset generated with the `synthpop` package, based on true data on daily number of hospital admissions for respiratory causes and ambient air NO_2 concentrations in the city of Barcelona (Spain) for years 2006-2015.

Examples

```
# time series:
opar <- par(no.readonly = TRUE)      # make a copy of current par settings
par(las = 1, mfrow = c(3, 1))
with(rhospno2, plot(date, hresp, type = "l", lwd = 0.5))
with(rhospno2, plot(date, no2, type = "l", lwd = 0.5))
with(rhospno2, plot(date, temp, type = "l", lwd = 0.5))
par(opar)                            # restore original par settings
```

`simulatedlnm` *simulatedllm generic.*

Description

Simulation procedure internally called by the `collindlnm` function for a given hypothetical effect.

Usage

```
simulatedlnm(x)

## S3 method for class 'linearlme'
simulatedlnm(x)

## S3 method for class 'linearglm'
simulatedlnm(x)

## S3 method for class 'nonlinearglm'
simulatedlnm(x)
```

Arguments

x	an object internally generated by the `collindlnm` function.

Value

a matrix with simulated results.

a matrix with simulated results.

a matrix with simulated results.

a matrix with simulated results.

Methods (by class)

- `linearlme`: case of a hypothetical linear effect in a model of class `"lme"`.
- `linearglm`: case of a hypothetical linear effect in a model of class `"glm"`.
- `nonlinearglm`: case of a hypothetical non-linear effect in a model of class `"glm"`.

Index