Package ‘dlsem’

January 10, 2019

Type Package

Title Distributed-Lag Linear Structural Equation Models

Version 2.4.1

Date 2019-01-11

Author Alessandro Magrini

Maintainer Alessandro Magrini <alessandro.magrini@unifi.it>

Description Inference functionalities for distributed-lag linear structural equation models (DLSEMs). DLSEMs are Markovian structural causal models where each factor of the joint probability distribution is a distributed-lag linear regression model (Magrini, 2018) <doi:10.2478/bile-2018-0012>. DLSEMs account for temporal delays in the dependence relationships among the variables and allow to perform dynamic causal inference by assessing causal effects at different time lags. Endpoint-constrained quadratic, quadratic decreasing and gamma lag shapes are available.

Depends graph, Rgraphviz

Imports methods

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2019-01-10 14:00:21 UTC

R topics documented:

`dlsem-package` *Distributed-lag linear structural equation models*

Description

Inference functionalities for distributed-lag linear structural equation models (DLSEMs). DLSEMs are Markovian structural causal models where each factor of the joint probability distribution is a distributed-lag linear regression model (Magrini, 2018). DLSEMs account for temporal delays in the dependence relationships among the variables and allow to perform dynamic causal inference by assessing causal effects at different time lags. Endpoint-constrained quadratic, quadratic decreasing and gamma lag shapes are available. The main functions of the package are:

- dlsem, to perform parameter estimation;
- causalEff, to compute all the pathwise causal lag shapes and the overall one connecting two or more variables;
- lagPlot, to display a pathwise or an overall causal lag shape.

Details

Package:	dlsem
Type:	Package
Version:	2.4.1
Date:	2019-01-11
License:	GPL-2

Author(s)

Alessandro Magrini <alessandro.magrini@unifi.it>

References

A. Magrini (2018). Linear Markovian Models for Lag Exposure Assessment. Biometrical Letters, 55(2): 179-195. DOI: 10.2478/bile-2018-0012.

`agres` *European agricultural data*

Description

Data on research expenditure, research activity, productivity and impacts of Agriculture from 1990 to 2014 in EU 15 countries. Source: EUROSTAT, Faostat, OECD, USDA.

Usage

```
data(agres)
```

Format

A data frame with a total of 350 observations on the following 20 variables:

`COUNTRY` Code of the country. Belgium and Luxembourg are merged together with code `BL`.

`YEAR` Year.

`GDP` Gross domestic product (million euro PPS).

`EMPL_AGR` Persons employed in Agriculture (count).

`UAA` Utilized agricultural area (hectares).

`PATENT_OTHER` Mechanical, chemical and environment-related patent applications (count).

`GBAORD_AGR` Government research expenditure (million euro PPS).

`BERD_AGR` Business enterprise research expenditure (million euro PPS).

`RD_EDU_AGR` Agricultural researchers with tertiary education (count).

`EU_PRO_AGR` Projects funded by the European Union (count).

`PATENT_AGR` Agricultural patent applications (count).

`TFPC` Total factor productivity (index 2005=100).

`EPC` Environmental performance (index 2005=100).

`GVA_AGR` Gross value added of Agriculture (international dollars).

`PPI_AGR` Producer price of agricultural output (index 2005=100).

`ENTR_INCOME_AGR` Net entrepreneurial income of Agriculture (index 2005=100).

`ENERGY_RENEW` Share of renewable energy in gross final energy consumption (index 2005=100).

`INCOME_RURAL` Mean familiar income in rural areas (international dollars).

`UNEMPL_RURAL` Unemployment rate in rural areas (index 2005=100).

`HEALTH_RURAL` Quota of persons with at least a good health status in rural areas (index 2005=100).

`as.graphNEL` *Conversion into the* `graphNEL` *class*

Description

An object of class `dlsem` is converted into an object of class `graphNEL`.

Usage

```
as.graphNEL(x, conf = 0.95, use.ns = FALSE)
```

Arguments

`x`	An object of class `dlsem`.
`conf`	The confidence level for each edge: only eadges with statistically significant causal effect at such level are considered. Default is 0.95.
`use.ns`	A logical value indicating whether edges without statistically significant causal effect (at level `conf`) should be considered or not. If `FALSE` (the default), they will be ignored.

Value

An object of class `graphNEL`.

See Also

dlsem.

Examples

```
data(industry)
indus.code <- list(
  Consum~quec.lag(Job,0,5),
  Pollution~quec.lag(Job,1,8)+quec.lag(Consum,1,6)
  )
indus.mod <- dlsem(indus.code,group="Region",exogenous=c("Population","GDP"),data=industry,
  log=TRUE)
as.graphNEL(indus.mod)
```

`auto.lagPlot` *Automated plot of lag shapes*

Description

All the single-edge pathwise causal lag shapes are saved as pdf files.

Usage

```
auto.lagPlot(x, cumul = FALSE, conf = 0.95, plotDir = NULL)
```

Arguments

x	An object of class `dlsem`.
cumul	Logical. If `TRUE`, cumulative causal effects are displayed. Default is `FALSE`.
conf	The confidence level for each plot. Default is 0.95.
plotDir	The directory where to save the plots. If `NULL` (the default), plots will be saved in the current working directory.

See Also

dlsem; lagPlot.

Examples

```
data(industry)
indus.code <- list(
  Consum~quec.lag(Job,0,5),
  Pollution~quec.lag(Job,1,8)+quec.lag(Consum,1,6)
  )
indus.mod <- dlsem(indus.code,group="Region",exogenous=c("Population","GDP"),data=industry,
  log=TRUE)
auto.lagPlot(indus.mod,plotDir=getwd())
```

`causalEff` *Assessment of dynamic causal effects*

Description

All the pathwise causal lag shapes and the overall one connecting two or more variables are computed.

Usage

```
causalEff(x, from = NULL, to = NULL, lag = NULL, cumul = FALSE, conf = 0.95,
  use.ns = FALSE)
```

Arguments

- `x` — An object of class `dlsem`.
- `from` — The name of the starting variable, or a vector containing the names of starting variables, which must be endogenous variables.
- `to` — The name of the ending variable, which must be an endogenous variable.
- `lag` — A non-negative integer or a vector of non-negative integers indicating the time lags to be considered. If `NULL`, the whole lag shapes will be considered.
- `cumul` — Logical. If `TRUE`, cumulative causal effects are returned. Default is `FALSE`.
- `conf` — The confidence level. Default is 0.95.
- `use.ns` — A logical value indicating whether edges without statistically significant causal effect (at level `conf`) should be considered or not. If `FALSE` (the default), they will be ignored.

Details

A *pathwise causal lag shape* is the set of causal effects associated to a path at different time lags. An *overall causal lag shape* is the set of overall causal effects of a variable on another one at different time lags.

Note that, due to the properties of the multiple linear regression model, causal effects are net of the influence of the group factor and exogenous variables.

Value

A list containing several matrices including point estimates, standard errors and asymptotic confidence intervals (at level `conf`) for all the pathwise causal lag shapes and the overall one connecting the starting variables to the ending variable.

Note

Value `NULL` is returned if one of the following occurs: (i) no significant path at confidence level `conf` exists connecting the starting variables to the ending variable; (ii) the requested path does not exist or is not significant at confidence level `conf`. Note that the edges between the starting variables and their respective parents are deleted as a consequence of intervention. See Magrini (2018) for technical details on causal inference in distributed-lag linear structural equation models.

References

A. Magrini (2018). Linear Markovian Models for Lag Exposure Assessment. Biometrical Letters, 55(2): 179-195. DOI: 10.2478/bile-2018-0012.

See Also

dlsem; lagPlot.

Examples

```
data(industry)
indus.code <- list(
  Consum~quec.lag(Job,0,5),
  Pollution~quec.lag(Job,1,8)+quec.lag(Consum,1,6)
  )
indus.mod <- dlsem(indus.code,group="Region",exogenous=c("Population","GDP"),data=industry,
  log=TRUE)
causalEff(indus.mod,from="Job",to="Pollution",lag=c(0,5,10,15),cumul=TRUE)
```

---

dlsem *Parameter estimation*

Description

Parameter estimation is performed for a distributed-lag linear structural equation model. A single group factor may be taken into account.

Usage

```
dlsem(model.code, group = NULL, time = NULL, exogenous = NULL, data,
  log = FALSE, hac = FALSE, diff.options = list(combine="choi",k=0,maxdiff=3),
  imput.options = list(tol=0.0001,maxiter=500,recons=TRUE,no.imput=NULL),
  global.control = NULL, local.control = NULL, quiet = FALSE)
```

Arguments

`model.code`	A list of objects of class `formula`, each describing a single regression model. See `Details`.
`group`	The name of the group factor (optional). If `NULL`, no groups are considered.
`time`	The name of the time factor (optional). If `time` is `NULL` and `group` is not `NULL`, data are assumed to be temporally ordered within each group. If both `time` and `group` are `NULL`, data are assumed to be temporally ordered.
`exogenous`	The name of exogenous variables (optional). Exogenous variables may be either quantitative or qualitative and must not appear in the model code.
`data`	An object of class `data.frame` containing data.
`hac`	Logical. If `TRUE`, heteroskedasticity and autocorrelation consistent estimation of standard errors (Newey & West, 1978) is applied. Default is `FALSE`.
`log`	Logical. If `TRUE`, logarithmic transformation is applied to all strictly positive quantitative variables. Default is `FALSE`.
`diff.options`	A list containing options for the differentiation. The list may consist of any number of components among the following: • `combine`: the method to combine p-values of the Augmented Dickey-Fuller test for different groups, that may be either `"choi"` or `"demetrescu"` (see unirootTest) Ignored if `group` is `NULL`. Default is `"choi"`;

- k: The lag order of the Augmented Dickey-Fuller test. If k=0 (the default), the standard Dickey-Fuller test is performed. If k=NULL, the lag order is taken as the cubic root of the sample size;
- maxdiff: the maximum differentiation order to apply. If maxdiff=0, differentiation will not be applied. Default is 3.

imput.options
: A list containing options for the imputation of missing values through the Expectation-Maximization algorithm (Dempster et al., 1977). The list may consist of any number of components among the following:

- tol: the tolerance threshold. Default is 0.0001;
- maxiter: the maximum number of iterations. Default is 500. If maxiter=0, imputation will not be performed;
- recons: a logical value indicating whether univariate reconstruction of time series should be performed before applying differentiation. Reconstruction is performed using cubic splines and is limited to the observed range of each time series (no extrapolation). It may be useful to avoid further missing values due to differentiation in time series with contiguous missing values. Default is TRUE.
- no.input: the name of variables to which reconstruction and imputation should not be applied.

global.control
: A list containing global options for the estimation. The list may consist of any number of components among the following:

- adapt: a logical value indicating whether adaptation of lag shapes should be performed for all regression models. Default is FALSE;
- min.gestation: the minimum gestation lag for all covariates. If not provided, it is assumed to be equal to 0. Ignored if adapt=FALSE;
- max.gestation: the maximum gestation lag for all covariates. If not provided, it is assumed to be equal to max.lead (see below). Ignored if adapt=FALSE;
- max.lead: the maximum lead lag for all covariates. If not provided, it is computed accordingly to the sample size. Ignored if adapt=FALSE;
- min.width: the minimum lag width for all covariates. It cannot be greater than max.lead. If not provided, it is assumed to be 0. Ignored if adapt=FALSE;
- sign: the sign of parameter θ_j (either '+' for positive or '-' for negative) for all covariates. If not provided, adaptation will disregard the sign of parameter θ_j. Ignored if adapt=FALSE;
- selection: the criterion to be used for the adaptation of lag shapes, that may be one among 'bic' to minimise the Bayesian Information Criterion (Schwarz, 1978) and 'aic' to minimise the Akaike Information Criterion (Akaike, 1974). Default is 'bic'.

local.control
: A list containing variable-specific options for the estimation. These options prevail on the ones contained in global.control. See Details.

quiet
: Logical. If TRUE, messages on the estimation progress are suppressed. Deafult is FALSE.

Details

Each regression model in a distributed-lag linear structural equation model has the form:

$$y_t = \beta_0 + \sum_{j=1}^{J} \sum_{l=0}^{L_j} \beta_{j,l}\, x_{j,t-l} + \epsilon_t$$

where y_t is the value of the response variable at time t, $x_{j,t-l}$ is the value of the j-th covariate at l time lags before t, and ϵ_t is the random error at time t uncorrelated with the covariates and with ϵ_{t-1}. The set $(\beta_{j,0}, \beta_{j,1}, \ldots, \beta_{j,L_j})$ is the lag shape of the j-th covariate.

For the endpoint-constrained quadratic lag shape:

$$\beta_{j,l} = \theta_j \left[-\frac{4}{(b_j - a_j + 2)^2} l^2 + \frac{4(a_j + b_j)}{(b_j - a_j + 2)^2} l - \frac{4(a_j - 1)(b_j + 1)}{(b_j - a_j + 2)^2} \right] \qquad a_j \leq l \leq b_j$$

(otherwise, $\beta_{j,l} = 0$).

For the quadratic decreasing lag shape:

$$\beta_{j,l} = \theta_j \frac{l^2 - 2(b_j + 1)l + (b_j + 1)^2}{(b_j - a_j + 1)^2} \qquad a_j \leq l \leq b_j$$

(otherwise, $\beta_{j,l} = 0$).

For the gamma lag shape:

$$\beta_{j,l} = \theta_j (l+1)^{\frac{\delta}{1-\delta}} \lambda_j^l \left[\left(\frac{\delta_j}{(\delta_j - 1)\log(\lambda_j)} \right)^{\frac{\delta_j}{1-\delta_j}} \lambda_j^{\frac{\delta_j}{(\delta_j - 1)\log(\lambda_j)} - 1} \right]^{-1}$$

$$0 < \delta_j < 1 \qquad 0 < \lambda_j < 1$$

Formulas cannot contain qualitative variables or interaction terms (no ':' or '*' symbols), and may contain the following operators for lag specification:

- `quec.lag`: quadratic (2nd order polynomial) lag shape with endpoint constraints;
- `qdec.lag`: quadratic (2nd order polynomial) decreasing lag shape;
- `gamm.lag`: gamma lag shape.

Each operator must have the following three arguments (provided within brackets and separated by commas):

1. the name of the covariate to which the lag is applied;
2. parameter a_j (for 2nd order polynomial lag shapes), or parameter δ_j (for the gamma lag shape);
3. parameter b_j (for 2nd order polynomial lag shapes), or parameter λ_j (for the gamma lag shape);
4. the group factor (optional). If not provided and argument `group` is not `NULL`, this is found automatically.

For example, `quec.lag(X1,3,15)` indicates that a quadratic lag shape with endpoint constraints must be applied to variable X1 in the interval (3,15), and `gamm.lag(X1,0.75,0.8)` indicates that a gamma lag shape with $\delta_j = 0.75$ and $\lambda_j = 0.8$ must be applied to variable X1. See Magrini (2018) for more details.

The formula of regression models with no endogenous covariates may be omitted from argument `model.code`. The group factor and exogenous variables must not appear in any formula.

Argument `local.control` must be a named list containing one or more among the following components:

- `adapt`: a named vector of logical values, where each component must have the name of one endogenous variable and indicate whether adaptation of lag shapes should be performed for the regression model of that variable.
- `min.gestation`: a named list. Each component of the list must have the name of one endogenous variable and be a named vector. Each component of the named vector must have the name of one covariate in the regression model of the endogenous variable above and include the minimum gestation lag for its lag shape.
- `max.gestation`: the same as `min.gestation`, with the exception that the named vector must include the maximum gestation lag.
- `max.lead`: the same as `min.gestation`, with the exception that the named vector must include the maximum lead lag.
- `min.width`: the same as `min.gestation`, with the exception that the named vector must include the minimum lag width.
- `sign`: the same as `min.gestation`, with the exception that the named vector must include the lag sign (either `'+'` for positive or `'-'` for negative). Local control options have no default values, and global ones are applied in their absence. If some local control options conflict with global ones, only the former are applied.

Imputation of missing values is performed after eventual logarithmic transformation and differentiation, assuming group-specific means and time-invariant covariance matrix.

Qualitative variables cannot contain missing values.

After differentiation, each quantitative variable must have at least a total of 3 observed values and, if the group factor is specified, at least one observed value per group.

Value

An object of class `dlsem`, with the following components:

`estimate`	A list of objects of class `lm`, one for each regression model.
`call`	A list containing the call for each regression model after eventual adaptation of lag shapes.
`exogenous`	The names of exogenous variables.
`group`	The name of the group factor. `NULL` is returned if `group=NULL`.
`log`	The value provided to argument `log`.
`ndiff`	The order of differentiation.
`diff.options`	Options used for the differentiation.

`imput.options`	Options used for the imputation of missing values.
`selection`	The criterion used for the adaptation of lag shapes.
`adaptation`	Variable-specific options used for the adaptation of lag shapes.
`Rsq`	The coefficient of determination (R-squared) for each regression model and the overall one.
`data.orig`	The dataset provided to argument `data`.
`data.used`	Data used in the estimation, that is after eventual logarithmic transformation and differentiation.

S3 methods available for class `dlsem` are:

`print` provides essential information on the model.

`summary` shows summaries of estimation.

`plot` displays the directed acyclic graph (DAG) of the model including only the endogenous variables. Option `conf` controls the confidence level (default is 0.95), while option `style` controls the style of the plot:

- `style=2` (the default): each edge is coloured with respect to the sign of the estimated causal effect (green: positive, red: negative, light grey: not statistically significant);
- `style=1`: edges with statistically significant causal effect are shown in black, otherwise they are shown in light grey;
- `style=0`: all edges are shown in black disregarding statistical significance of causal effects.

`nobs` returns the number of observations for each regression model.

`npar` returns the number of parameters for each regression model.

`coef` returns parameters $\theta_j (j = 1, \ldots, J)$ for each regression model.

`vcov` returns the covariance matrix of estimates for each regression model.

`logLik` returns the log-likelihood for each regression model.

`AIC` returns the Akaike Information Criterion for each regression model and the overall one.

`BIC` returns the Bayesian Information Criterion for each regression model and the overall one.

`fitted` returns fitted values.

`residuals` returns residuals.

`predict` returns predicted values. If k-th order differentiation was applied, predicted values represent variations (on the original scale if `log=FALSE`, percentage variations if `log=TRUE`) with respect to k time lags before. Optionally, a data frame from which to predict may be provided to argument `newdata`.

References

W. K. Newey, and K. D. West (1978). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), 703-708.

H. Akaike (1974). A New Look at the Statistical Identification Model. *IEEE Transactions on Automatic Control*, 19, 716-723. DOI: 10.1109/TAC.1974.1100705

A. P. Dempster, N. M. Laird, and D. B. Rubin (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, Series B, 39(1): 1-38.

G. Schwarz (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6, 461-464. DOI: 10.1214/aos/1176344136

See Also

unirootTest; causalEff.

Examples

```
data(industry)

# Estimation without adaptation of lag shapes
indus.code <- list(
  Consum~quec.lag(Job,0,5),
  Pollution~quec.lag(Job,1,8)+quec.lag(Consum,1,6)
  )
indus.mod <- dlsem(indus.code,group="Region",time="Year",exogenous=c("Population","GDP"),
  data=industry,log=TRUE)

# Adaptation of lag shapes (estimation takes some seconds more)
indus.global <- list(adapt=TRUE,max.gestation=3,max.lead=15,min.width=5,sign="+")
indus.local <- list()
## NOT RUN:
# indus.mod <- dlsem(indus.code,group="Region",time="Year",exogenous=c("Population","GDP"),
#   data=industry,global.control=indus.global,local.control=indus.local,log=TRUE)

# Summary of estimation
summary(indus.mod)

# DAG with coloured edges
plot(indus.mod)

# DAG disregarding statistical significance
plot(indus.mod,style=0)


### Comparison among alternative models

# Model 2: quadratic decreasing lag shapes
indus.code_2 <- list(
  Job ~ 1,
  Consum~qdec.lag(Job,0,15),
  Pollution~qdec.lag(Job,0,15)+qdec.lag(Consum,0,15)
```

```
  )
## NOT RUN:
# indus.mod_2 <- dlsem(indus.code_2,group="Region",time="Year",exogenous=c("Population","GDP"),
#   data=industry,global.control=indus.global,local.control=indus.local,log=TRUE)

# Model 3: gamma lag shapes
indus.code_3 <- list(
  Job ~ 1,
  Consum~gamm.lag(Job,0.5,0.5),
  Pollution~gamm.lag(Job,0.5,0.5)+gamm.lag(Consum,0.5,0.5)
  )
## NOT RUN:
# indus.mod_3 <- dlsem(indus.code_3,group="Region",time="Year",exogenous=c("Population","GDP"),
#   data=industry,global.control=indus.global,local.control=indus.local,log=TRUE)

# comparison of the three models
## NOT RUN:
# lapply(list(QUEC=indus.mod,QDEC=indus.mod_2,GAMMA=indus.mod_3),BIC)


### A more complex model

data(agres)

# Qualitative exogenous variable
agres$POLICY <- factor(1*(agres$YEAR>=2005))
levels(agres$POLICY) <- c("no","yes")

# Causal levels
context.var <- c("GDP","EMPL_AGR","UAA","PATENT_OTHER","POLICY")
investment.var <- c("GBAORD_AGR","BERD_AGR")
research.var <- c("RD_EDU_AGR","PATENT_AGR")
impact.var <-  c("ENTR_INCOME_AGR","PPI_AGR")

# Constraints on lag shapes
agres.global <- list(adapt=TRUE,max.gestation=3,min.width=5,max.lead=20,sign="+")
auxcon1 <- rep(15,length(investment.var))
names(auxcon1) <- investment.var
auxcon2 <- rep("-",length(investment.var)+length(research.var))
names(auxcon2) <- c(investment.var,research.var)
agres.local <- list(max.lead=list(RD_EDU_AGR=auxcon1,PATENT_AGR=auxcon1),
  sign=list(PPI_AGR=auxcon2))

# Endpoint-constrained quadratic lag shapes (estimation takes a couple of minutes)
auxcode <- c(paste(investment.var,"~1",sep=""),
  paste(research.var,"~",paste("quec.lag(",investment.var,",0,20)",
    collapse="+",sep=""),sep=""),
  paste(impact.var,"~",paste("quec.lag(",c(investment.var,research.var),",0,20)",
    collapse="+",sep=""),sep=""))
agres.code <- list()
for(i in 1:length(auxcode)) {
  agres.code[[i]] <- formula(auxcode[i])
  }
```

```
## NOT RUN:
# agres.mod <- dlsem(agres.code,group="COUNTRY",time="YEAR",exogenous=context.var,
#   data=agres,global.control=agres.global,local.control=agres.local,log=TRUE)
# summary(agres.mod)

## Gamma lag shapes (estimation takes some minutes)
auxcode_2 <- c(paste(investment.var,"~1",sep=""),
  paste(research.var,"~",paste("gamm.lag(",investment.var,",0.5,0.5)",
    collapse="+",sep=""),sep=""),
  paste(impact.var,"~",paste("gamm.lag(",c(investment.var,research.var),",0.5,0.5)",
    collapse="+",sep=""),sep=""))
agres.code_2 <- list()
for(i in 1:length(auxcode_2)) {
  agres.code_2[[i]] <- formula(auxcode_2[i])
  }
## NOT RUN:
# agres.mod_2 <- dlsem(agres.code_2,group="COUNTRY",time="YEAR",exogenous=context.var,
#  data=agres,global.control=agres.global,local.control=agres.local,log=T)
# summary(agres.mod_2)
# lapply(list(QUEC=agres.mod,GAMMA=agres.mod_2),BIC)
```

`industry` *Industrial development*

Description

Simulated data on industrial development from 1983 to 2014 in 10 imaginary regions.

Usage

```
data(industry)
```

Format

A data frame with a total of 320 observations on the following 7 variables:

`Region` ID of the region.

`Year` Year.

`Population` Population (number of inhabitants).

`GDP` Gross domestic product (1000 international dollars).

`Job` Employment in industry (Index, 1980=100).

`Consum` Private consumption (Index, 1980=100).

`Pollution` Greenhouse gas emissions (tonnes of CO2 equivalent).

isIndep *Conditional independence check*

Description

Conditional independence between two variables is checked using the d-separation criterion (Pearl, 2000, page 16 and following).

Usage

```
isIndep(x, var1 = NULL, var2 = NULL, given = NULL, conf = 0.95, use.ns = FALSE)
```

Arguments

`x`	An object of class `dlsem`.
`var1`	The name of the first variable.
`var2`	The name of the second variable.
`given`	A vector containing the names of conditioning variables. If `NULL`, marginal independence is checked.
`conf`	The confidence level for each edge: only eadges with statistically significant causal effect at such level are considered. Default is 0.95.
`use.ns`	A logical value indicating whether edges without statistically significant causal effect (at level `conf`) should be considered or not. If `FALSE` (the default), they will be ignored.

Value

Logical

Note

Conditional independence is checked statically, that is the whole history of conditioning variables is supposed to be known.

The result is unchanged if arguments `var1` and `var2` are switched.

Dependence is a necessary but not sufficient condition for causation: see the discussion in Pearl (2000).

References

J. Pearl (2000). Causality: Models, Reasoning, and Inference. Cambridge University Press. Cambridge, UK. ISBN: 978-0-521-89560-6

See Also

dlsem.

Examples

```
data(industry)
indus.code <- list(
  Consum~quec.lag(Job,0,5),
  Pollution~quec.lag(Job,1,8)+quec.lag(Consum,1,6)
  )
indus.mod <- dlsem(indus.code,group="Region",exogenous=c("Population","GDP"),data=industry,
  log=TRUE)
isIndep(indus.mod,"Job","Pollution",given=c("Consum"))
```

`lagPlot` *Plot lag shapes*

Description

A pathwise or an overall causal lag shape is displayed.

Usage

```
lagPlot(x, from = NULL, to = NULL, path = NULL, maxlag = NULL, cumul = FALSE,
  conf = 0.95, use.ns = FALSE, ylim = NULL, title = NULL)
```

Arguments

`x`	An object of class `dlsem`.
`from, to, path`	To display the overall causal lag shape of a variable to another one, their names must be provided to arguments `from` and `to`, respectively. To display a pathwise causal lag shape, the name of the path, indicated as a string made of the names of the variables in the path separated by '*', must be provided to argument `path`. Argument `path` will be ignored if both `from` and `to` are not `NULL`.
`maxlag`	The maximum lag displayed (optional).
`cumul`	Logical. If `TRUE`, cumulative causal effects are returned. Default is `FALSE`.
`conf`	The confidence level for each edge: only statistically significant edges at such level are considered. Default is 0.95.
`use.ns`	A logical value indicating whether not statistically significant edges (at level `conf`) should be considered or not. If `FALSE` (the default), they will be ignored.
`ylim`	A vector of two numerical values indicating the limits of the y axis (optional). If `NULL`, the limits of the y axis are computed automatically.
`title`	The title of the plot (optional). If `NULL`, a default title is used.

Note

Value `NULL` is returned if one of the following occurs: (i) no significant path at confidence level `conf` exists connecting the starting variables to the ending variable; (ii) the requested path does not exist or is not significant at confidence level `conf`.

See Also

dlsem; causalEff; lagPlot.

Examples

```
data(industry)
indus.code <- list(
  Consum~quec.lag(Job,0,5),
  Pollution~quec.lag(Job,1,8)+quec.lag(Consum,1,6)
  )
indus.mod <- dlsem(indus.code,group="Region",exogenous=c("Population","GDP"),data=industry,
  log=TRUE)

# the lag shape of the causal effect associated to specific paths
lagPlot(indus.mod,path="Job*Pollution")
lagPlot(indus.mod,path="Job*Consum*Pollution")

# the lag shape of an overall causal effect
lagPlot(indus.mod,from="Job",to="Pollution")
```

---

uncons.lag *Lag shape transformation for model formulas*

Description

Lag shape transformation to be used in model formulas.

Usage

```
quec.lag(x, a, b, x.group = NULL)
qdec.lag(x, a, b, x.group = NULL)
gamm.lag(x, delta, lambda, x.group = NULL)
uncons.lag(x, a, b, x.group = NULL)
almon.lag(x, a, b, p, x.group = NULL)
```

Arguments

`x`	The name of the variable.
`a`	The gestation lag.
`b`	The lead lag.
`p`	The polynomial order for the Almon's lag shape (only for `almon.lag`).
`delta, lambda`	Delta and lambda parameters for the gamma lag shape (only for `gamm.lag`).
`x.group`	The name of the group factor (optional).

Examples

```
data(industry)
# example in linear regression
m0 <- lm(Consum ~ -1+Region+quec.lag(Job,0,5,x.group=Region), data=industry)
m1 <- lm(Consum ~ -1+Region+qdec.lag(Job,0,5,x.group=Region), data=industry)
m2 <- lm(Consum ~ -1+Region+gamm.lag(Job,0.5,0.5,x.group=Region), data=industry)
m3 <- lm(Consum ~ -1+Region+uncons.lag(Job,0,3,x.group=Region), data=industry)
m4 <- lm(Consum ~ -1+Region+almon.lag(Job,0,3,2,x.group=Region), data=industry)
```

---

uniroottest *Unit root test*

Description

The Augmented Dickey-Fuller test is performed on a set of quantitative variables. A single group factor may be taken into account.

Usage

```
unirootTest(x = NULL, group = NULL, time = NULL, data, combine = "choi", k = 0,
  log = FALSE)
```

Arguments

x	A vector including the name of the quantitative variables to be tested. If NULL (the default), all the quantitative variables contained in data will be tested.
group	The name of the group factor (optional). If NULL, no groups are considered.
time	The name of the time factor (optional). If time is NULL and group is not NULL, data are assumed to be temporally ordered within each group. If both time and group are NULL, data are assumed to be temporally ordered.
data	An object of class data.frame containing the variables to be tested, the group factor if group is not NULL, and the time factor if time is not NULL.
combine	The method to combine p-values of different groups, that may be either "choi" (Choi, 2001) or "demetrescu" (Demetrescu, 2006). Ignored if group is NULL. Default is "choi".
k	The lag order to calculate the statistic of the Augmented Dickey-Fuller test. If k=0 (the default), the standard Dickey-Fuller test is performed. If k=NULL, the lag order is taken as the cubic root of the sample size.
log	Logical. If TRUE, logarithmic transformation is applied to all strictly positive quantitative variables. Default is FALSE.

Value

An object of class unirootTest, consisting of a list with one component for each variable tested. Each list contains the following components:

statistic	The value of the test statistic.
n	The total number of observations if group is NULL, otherwise the number of observations per group.
z.value	The z-value of the test.
p.value	The p-value of the test.

Note

The null hypothesis of the Augmented Dickey-Fuller test is the presence of a unit root.

References

I. Choi (2001). Unit Root Tests for Panel Data. *Journal of International Money and Finance*, 20, 249-272. DOI: 10.1016/S0261-5606(00)00048-6

M. Demetrescu, U. Hassler, and A. Tarcolea (2006). Combining Significance of Correlated Statistics with Application to Panel Data. *Oxford Bulletin of Economics and Statistics*, 68(5), 647-663. DOI: 10.1111/j.1468-0084.2006.00181.x

D. A. Dickey, and W. A. Fuller (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49: 1057-1072. DOI: 10.2307/1912517

Examples

```
data(industry)
unirootTest(c("Job","Consum","GDP"),group="Region",time="Year",data=industry)
```

---

vcovHAC *Heteroskedasticty and autocorrelation consistent covariance matrix*

Description

The heteroskedasticty and autocorrelation consistent (HAC) covariance matrix of least square estimates (Newey & West, 1978) is computed. A single group factor may be taken into account.

Usage

```
vcovHAC(x, group = NULL)
```

Arguments

x	An object of class lm or dlsem.
group	The name of the group factor (optional). If NULL, no groups are considered.

Value

A matrix if x is of class lm, or, if x is of class dlsem, a list of matrices, one for each regression model. Each matrix has the attribute max.lag, indicating the maximum lag of autocorrelation, automatically computed based on fit to data. If group is not NULL, this is computed within each group.

Note

Residuals are assumed to be temporally ordered within each group.

References

W. K. Newey, and K. D. West (1978). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), 703-708.

Examples

```
data(industry)
m0 <- lm(Consum ~ -1+Region+quec.lag(Job,0,5,x.group=Region), data=industry)
# traditional covariance matrix
S <- vcov(m0)
# HAC covariance matrix
S_hac <- vcovHAC(m0,group="Region")
# comparison of standard errors
cbind(traditional=sqrt(diag(S)),hac=sqrt(diag(S_hac)))
```

Index