Package 'dlsem'

July 22, 2025

Type Package

Title Distributed-Lag Linear Structural Equation Models

Version 2.4.6

Date 2020-03-23

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 with constrained lag shapes (Magrini, 2018 <doi:10.2478/bile-2018-0012>; Magrini et al., 2019 <doi:10.1007/s11135-019-00855-z>). DLSEMs account for temporal delays in the dependence relationships among the variables through a single parameter per covariate, thus allowing to perform dynamic causal inference in a feasible fashion. Endpoint-constrained quadratic, quadratic decreasing, linearly decreasing and gamma lag shapes are available.

Depends R (>= 3.5.0), graph, Rgraphviz

Imports methods

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2020-04-16 16:10:11 UTC

Contents

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 with constrained lag shapes (Magrini, 2018; Magrini et. al, 2019; Magrini, 2020). DLSEMs account for temporal delays in the dependence relationships among the variables through a single parameter per covariate, thus allowing to perform dynamic causal inference in a feasible fashion. Endpoint-constrained quadratic ('ecq'), quadratic decreasing ('qd'), linearly decreasing ('ld') and gamma ('gam') 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:	2.4.6
Date:	2020-03-22
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.

A. Magrini, F. Bartolini, A. Coli, B. Pacini (2019). A structural equation model to assess the impact of agricultural research expenditure on multiple dimensions. Quality and Quantity, 53(4): 2063-2080. DOI: 10.1007/s11135-019-00855-z

A. Magrini (2020). A family of theory-based lag shapes for distributed-lag linear regression. To be appeared on *Italian Journal of Applied Statistics*.

---

agres	*European agricultural data*

Description

Data on research expenditure, research activity, productivity and impacts of Agriculture from 1990 to 2014 in EU 15 countries.

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). Source: Eurostat.

EMPL_AGR Persons employed in Agriculture (count). Source: Eurostat.

UAA Utilized agricultural area (hectares). Source: Eurostat.

PATENT_OTHER Mechanical, chemical and environment-related patent applications (count). Source: OECD.

GBAORD_AGR Government research expenditure (million euro PPS). Source: OECD.

BERD_AGR Business enterprise research expenditure (million euro PPS). Source: Eurostat.

RD_EDU_AGR Agricultural researchers with tertiary education (count). Source: OECD.

PATENT_AGR Agricultural patent applications (count). Source: OECD.

TFP_AGR Total factor productivity (index 2005=100). Source: USDA.

GVA_AGR Gross value added of Agriculture (international dollars). Source: Faostat.

PPI_AGR Producer price of agricultural output (index 2005=100). Source: Faostat.

ENTR_INCOME_AGR Net entrepreneurial income of Agriculture (index 2005=100). Source: Eurostat.

GHG_AGR Greenhouse gas emissions due to Agriculture (gigagrams CO_2 equivalent). Source: Faostat.

NBAL_AGR Nitrogen gross nutrient balance (kg of nutrient per hectare). Source: Eurostat.

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

UNEMPL_RURAL Unemployment rate 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 edges with statistically significant causal effect at such confidence 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~ecq(Job,0,5),
  Pollution~ecq(Job,1,8)+ecq(Consum,1,7)
  )
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~ecq(Job,0,5),
  Pollution~ecq(Job,1,8)+ecq(Consum,1,7)
  )
indus.mod <- dlsem(indus.code,group="Region",exogenous=c("Population","GDP"),data=industry,
  log=TRUE)
## NOT RUN:
# auto.lagPlot(indus.mod,plotDir=getwd())
```

---

autoCode	*Automated model code*

Description

Given a set of variable names, a model code including all the possible edges is built.

Usage

```
autoCode(var.names, lag.type = "ecq")
```

Arguments

var.names	A vector containing the names of the variables, which must be at least of length 2.
lag.type	The type of lag shape, which will be applied to all variables and must be one among 'ecq', 'qd', 'ld', 'gam' and 'none'. Default is 'ecq'.

Value

A list of formulas to be passed as argument model.code in dlsem().

See Also

[dlsem](#).

Examples

```
autoCode(c("Job","Consum","Pollution"),lag.type="ecq")
```

---

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~ecq(Job,0,5),
  Pollution~ecq(Job,1,8)+ecq(Consum,1,7)
  )
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)
```

---

compareModels   *Comparison among different distributed-lag linear structural equation models*

Description

Several competing distributed-lag linear structural equation models are compared based on information criteria.

Usage

```
compareModels(x)
```

Arguments

x   A list of 2 or more objects of class `dlsem` estimated on the same data.

Value

A data.frame with one record for each model in `x` on the following quantities: log-likelihood, number of parameters, Akaike Information Criterion (AIC), Bayesian Information criterion (BIC).

Note

In order to keep the sample size constant, only the non-missing residuals across all the models are considered (see Magrini, 2020, for details).

References

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. Magrini (2020). A family of theory-based lag shapes for distributed-lag linear regression. To be appeared on *Italian Journal of Applied Statistics*.

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

See Also

[dlsem](#).

Examples

```
data(industry)

# model with endpoint-contrained quadratic lag shapes
indus.code <- list(
  Consum~ecq(Job,0,5),
  Pollution~ecq(Job,1,8)+ecq(Consum,1,7)
```

```
  )
indus.mod <- dlsem(indus.code,group="Region",exogenous=c("Population","GDP"),data=industry,
  log=TRUE)

# model with gamma lag shapes
indus.code_2 <- list(
  Consum~gam(Job,0.85,0.2),
  Pollution~gam(Job,0.95,0.05)+gam(Consum,0.9,0.15)
  )
indus.mod_2 <- dlsem(indus.code_2,group="Region",exogenous=c("Population","GDP"),data=industry,
  log=TRUE)

compareModels(list(indus.mod,indus.mod_2))
```

---

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, hac = TRUE,
  gamma.by = 0.05, global.control = NULL, local.control = NULL, log = FALSE,
  diff.options = list(test=NULL, maxdiff=2, ndiff=NULL),
 imput.options = list(tol=0.0001, maxiter=500, maxlag=2, no.imput=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 variable indicating the date time (optional). This variable must be either a numeric identificative or a date in format '%Y/%m/%d','%d/%m/%Y', or '%Y-%m-%d'. 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 (HAC) estimation of standard errors by Newey & West (1978) is applied, otherwise OLS standard errors are used. Default is `TRUE`.

gamma.by   A real number between 0 and 1 defining the resolution of the grid for the adaptation of gamma lag shapes. Adaptation is more precise with values near 0, but it will also take more time. Default is 0.05.

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 for a gamma lag shape or 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 for a gamma lag shape or 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. Note that the lead lag for a gamma lag shape is determined numerically;
- 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 for a gamma lag shape or if adapt=FALSE;
- sign: the sign of parameter θ_i (either '+' for positive or '-' for negative) for all covariates. If not provided, adaptation will disregard the sign of parameter θ_i. Ignored if adapt=FALSE.

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

log   Logical or a vector of characters. If a vector of characters is provided, logarithmic transformation is applied to strictly positive quantitative variables with name matching those characters. If TRUE, logarithmic transformation is applied to all strictly positive quantitative variables. Default is FALSE.

diff.options   A list containing options for differencing. The list may consist of any number of components among the following:

- test: the unit root test to apply, that may be either "kpss" for KPSS test or "adf" for ADF test (see [unirootTest](#)). If NULL (the default), the choice is for the KPSS test if the number of periods is less than 100, otherwise the ADF test is used.
- maxdiff: the maximum differencing order to apply. If maxdiff=0, differencing will not be applied. Default is 3;
- ndiff: the order of differencing to apply (without performing unit root test). If ndiff=NULL, differencing will be applied according to unit root test. Default is NULL.

Differencing is applied to remove unit roots, thus avoiding spurious regression (Granger \& Newbold, 1974). The same order of differencing will be applied to all quantitative variables.

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;
- `maxlag`: The maximum autoregressive order to consider in the imputation. Default is 3.
- `no.input`: the name of variables to which imputation should not be applied.

Only missing values of quantitative variables will be imputed. Qualitative variables cannot contain missing values.

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:

$$y_t = \beta_0 + \sum_{i=1}^{p} \sum_{l=0}^{L_i} \beta_{i,l}\, x_{i,t-l} + \epsilon_t$$

where y_t is the value of the response variable at time t, $x_{i,t-l}$ is the value of the i-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 $(\beta_{i,0}, \beta_{i,1}, \ldots, \beta_{i,L_i})$ is the lag shape of the i-th covariate. Currently available lag shapes are the endpoint-constrained quadratic lag shape:

$$\beta_{i,l} = \theta_i \left[-\frac{4}{(b_i - a_i + 2)^2} l^2 + \frac{4(a_i + b_i)}{(b_i - a_i + 2)^2} l - \frac{4(a_i - 1)(b_i + 1)}{(b_i - a_i + 2)^2} \right] \qquad a_i \le l \le b_i$$

(otherwise, $\beta_{i,l} = 0$); the quadratic decreasing lag shape:

$$\beta_{i,l} = \theta_i \frac{l^2 - 2(b_i + 1)l + (b_i + 1)^2}{(b_i - a_i + 1)^2} \qquad a_i \le l \le b_i$$

(otherwise, $\beta_{i,l} = 0$); the linearly decreasing lag shape:

$$\beta_{i,l} = \theta_i \frac{b_i + 1 - l}{b_i + 1 - a_i} \qquad a_i \le l \le b_i$$

(otherwise, $\beta_{i,l} = 0$); the gamma lag shape:

$$\beta_{i,l} = \theta_i (l+1)^{\frac{a_i}{1-a_i}} b_i^l \left[\left(\frac{a_i}{(a_i - 1)\log(b_i)} \right)^{\frac{a_i}{1-a_i}} b_i^{\frac{a_i}{(a_i - 1)\log(b_i)} - 1} \right]^{-1}$$

$$0 < a_i < 1 \qquad 0 < b_i < 1$$

See Magrini (2020) for details on these constrained lag shapes.

Formulas cannot contain neither qualitative variables or interaction terms (no ':' or '*' symbols), nor functions excepting the following operators for the specification of lag shapes:

- `ecq`: quadratic (2nd order polynomial) lag shape with endpoint constraints;
- `qd`: quadratic (2nd order polynomial) decreasing lag shape;

- ld: linearly decreasing lag shape;
- gam: 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_i;
3. parameter b_i;
4. the group factor (optional). If not provided and argument group is not NULL, this is found automatically.

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.

Value

An object of class dlsem, with the following components:

estimate	A list of objects of class lm, one for each regression model.
model.code	The model code, eventually after adaptation of the lag shapes.
call	A list containing the call for each regression, eventually after adaptation of lag shapes.
lag.par	A list containing the parameters of the lag shapes for each regression, eventually after adaptation.
exogenous	The names of exogenous variables.

group	The name of the group factor. NULL is returned if `group=NULL`.
time	The name of the variable indicating the date time. NULL is returned if `time=NULL`.
log	The name of the variables to which logarithmic transformation was applied.
ndiff	The order of differencing applied to the time series.
data	The data after eventual logarithmic transformation and differencing, which were actually used to estimate the model.
data.orig	The dataset provided to argument `data`.
fitted	The fitted values.
residuals	The residuals.
autocorr	The estimated order of the residual auto-correlation for each regression model.
hac	Logical indicating whether HAC estimation of standard errors is applied.
adaptation	Variable-specific options used for the adaptation of lag shapes.

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 the estimates of scale parameters for each regression model.
confint	returns the confidence intervals of scale parameters for each regression model. Argument `level` controls the confidence level (default is 0.95).
vcov	returns the covariance matrix of estimates for each regression model.
logLik	returns the log-likelihood for each regression model.
fitted	returns fitted values.
residuals	returns residuals.
predict	returns predicted values. Optionally, a data frame from which to predict may be provided to argument `newdata`.

References

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.

C. W. J. Granger, and P. Newbold (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2): 111-120.

A. Magrini (2019). A family of theory-based lag shapes for distributed-lag linear regression. To be appeared on *Italian Journal of Applied Statistics*.

W. K. Newey, and K. D. West (1978). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703-708.

See Also

unirootTest; causalEff; compareModels.

Examples

```
data(industry)

# Estimation without adaptation of lag shapes
indus.code <- list(
  Consum~ecq(Job,0,5),
  Pollution~ecq(Job,1,8)+ecq(Consum,1,7)
  )
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=5,max.lead=15,min.width=3,sign="+")
## NOT RUN:
# indus.mod <- dlsem(indus.code,group="Region",time="Year",exogenous=c("Population","GDP"),
#   data=industry,global.control=indus.global,log=TRUE)

# Summary of estimation
summary(indus.mod)$endogenous

# DAG with edges coloured according to the sign
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~qd(Job),
  Pollution~qd(Job)+qd(Consum)
  )
```

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

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

# Model 4: gamma lag shapes
indus.code_4 <- list(
  Job ~ 1,
  Consum~gam(Job),
  Pollution~gam(Job)+gam(Consum)
  )
## NOT RUN:
# indus.mod_4 <- dlsem(indus.code_4,group="Region",time="Year",exogenous=c("Population","GDP"),
#   data=industry,global.control=indus.global,log=TRUE)

# comparison of the three models
## NOT RUN:
# compareModels(list(indus.mod,indus.mod_2,indus.mod_3,indus.mod_4))


### 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("GVA_AGR","PPI_AGR")
agres.var <- c(context.var,investment.var,research.var,impact.var)

# Constraints on lag shapes
agres.global <- list(adapt=TRUE,max.gestation=5,max.lead=15,sign="+")
agres.local <- list(
  sign=list(
    PPI_AGR=c(GBAORD_AGR="-",BERD_AGR="-",RD_EDU_AGR="-",PATENT_AGR="-")
    )
  )

# Endpoint-constrained quadratic lag shapes (estimation takes a couple of minutes)
```

```
auxcode <- c(paste(investment.var,"~1",sep=""),
  paste(research.var,"~",paste("ecq(",investment.var,",,)",
    collapse="+",sep=""),sep=""),
  paste(impact.var,"~",paste("ecq(",c(investment.var,research.var),",,)",
    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)$endogenous

## Gamma lag shapes (estimation takes some minutes)
auxcode_2 <- c(paste(investment.var,"~1",sep=""),
  paste(research.var,"~",paste("gam(",investment.var,",,)",
    collapse="+",sep=""),sep=""),
  paste(impact.var,"~",paste("gam(",c(investment.var,research.var),",,)",
    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=TRUE)
# summary(agres.mod_2)$endogenous
# compareModels(list(agres.mod,agres.mod_2))
```

---

drawSample   *Sampling from a distributed-lag linear structural equation model*

Description

A future sample from a distributed-lag linear structural equation model is drawn.

Usage

```
drawSample(x, n)
```

Arguments

x	An object of class `dlsem`.
n	The sample size (temporal horizon).

Value

An object of class `data.frame`.

Note

Sampling is conditioned to the most recent observed value of all the variables.

For variables subdued to logarithmic transformation and/or differencing, the sampled values are in logarithmic scale and/or after differencing, as well.

If a group factor was specified for the model, a sample of size n is drawn for each group.

See Also

[dlsem](#).

Examples

```
data(industry)
indus.code <- list(
  Consum~ecq(Job,0,5),
  Pollution~ecq(Job,1,8)+ecq(Consum,1,7)
  )
indus.mod <- dlsem(indus.code,group="Region",time="Year",exogenous=c("Population","GDP"),
  data=industry,log=TRUE)
drawSample(indus.mod,10)
```

---

ecq   *Lag shape constructors*

Description

Lag shape constructors to be used in model formulas.

Usage

```
ecq(x, a, b, x.group = NULL, nlag = NULL)
qd(x, a, b, x.group = NULL, nlag = NULL)
ld(x, a, b, x.group = NULL, nlag = NULL)
gam(x, a, b, x.group = NULL, nlag = NULL)
```

Arguments

x	The name of the variable.
a, b	The shape parameters.
nlag	The number of lags considered. If NULL or more than 2/3 of the sample size, it is set equal to 2/3 of the sample size.
x.group	The name of the group factor (optional).

References

A. Magrini (2020). A family of theory-based lag shapes for distributed-lag linear regression. To be appeared on *Italian Journal of Applied Statistics*.

Examples

```
data(industry)
# example in linear regression
m1 <- lm(Consum ~ -1+Region+ecq(Job,0,5,x.group=Region), data=industry)
m2 <- lm(Consum ~ -1+Region+gam(Job,0.85,0.2,x.group=Region), data=industry)
```

---

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 CO_2 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 edges with statistically significant causal effect at such confidence 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~ecq(Job,0,5),
  Pollution~ecq(Job,1,8)+ecq(Consum,1,7)
  )
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 of 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 confidence 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; lagShapes; causalEff.

Examples

```
data(industry)
indus.code <- list(
  Consum~ecq(Job,0,5),
  Pollution~ecq(Job,1,8)+ecq(Consum,1,7)
  )
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")
```

---

lagShapes   *Estimated lag shapes*

Description

Estimated lag shapes and their standard errors are provided.

Usage

```
lagShapes(x, cumul = FALSE)
```

Arguments

x	An object of class dlsem.
cumul	Logical. If TRUE, cumulative causal effects are returned. Default is FALSE.

Value

A list of lists, one for each endogenous variable, each containing several matrices including the estimated lag shapes and their standard errors.

See Also

[dlsem](); [causalEff](); [lagPlot]().

Examples

```
data(industry)
indus.code <- list(
  Consum~ecq(Job,0,5),
  Pollution~ecq(Job,1,8)+ecq(Consum,1,7)
  )
indus.mod <- dlsem(indus.code,group="Region",exogenous=c("Population","GDP"),data=industry,
```

```
   log=TRUE)
lagShapes(indus.mod)
```

---

lmHAC   *Heteroskedasticty and autocorrelation consistent covariance matrix*

Description

The heteroskedasticty and autocorrelation consistent (HAC) covariance matrix of least square esti-
mates (Newey & West, 1978) is applied to an object of class `lm`. A single group factor may be taken
into account.

Usage

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

Arguments

x   An object of class `lm`.

group   The name of the group factor (optional). If NULL, no groups are considered.

Value

An object of class `hac` and `lm`. The HAC covariance matrix is stored into the component `vcov` of
the object, which is taken into account by the `summary` and the `vcov` methods. The HAC covari-
ance matrix has the attribute `max.lag`, indicating the maximum lag of autocorrelation, which is
automatically computed based on fit to data.

Note

If `group` is not NULL, the HAC covariance matrix is computed within each group. 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+ecq(Job,0,5,x.group=Region), data=industry)
summary(m0)
confint(m0)

m0_hac <- lmHAC(m0,group="Region")
summary(m0_hac)
confint(m0_hac)
```

---

residualPlot   *Plot for diagnostics of the residuals in a distributed-lag structural equation model.*

Description

Several different types of plot for diagnostics of the residuals in a distributed-lag structural equation model are displayed.

Usage

```
residualPlot(x, type = "fr")
```

Arguments

x   An object of class dlsem.

type   A character string indicating the type of plot, which should be one among 'fr' (fitted vs. residuals plot, the default), 'qq' (quantile-quantile plot of the residuals), 'ts' (time series plot of the residuals), 'ac' (auto-correlation plot of the residuals).

Note

If type is equal to 'ts' or 'ac' and a group factor was specified for x, the results are displayed as minimum, 1st quartile, median, 3rd quartile and maximum by group.

See Also

[dlsem](#).

Examples

```
data(industry)
indus.code <- list(
  Consum~ecq(Job,0,5),
  Pollution~ecq(Job,1,8)+ecq(Consum,1,7)
  )
indus.mod <- dlsem(indus.code,group="Region",exogenous=c("Population","GDP"),data=industry,
  log=TRUE)
residualPlot(indus.mod,type="fr")
residualPlot(indus.mod,type="qq")
residualPlot(indus.mod,type="ts")
residualPlot(indus.mod,type="ac")
```

---

unirootTest   *Unit root test*

Description

Unit root 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, test = NULL, 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). This variable must be either a numeric identificative or a date in format '%Y/%m/%d','%d/%m/%Y', or '%Y-%m-%d'. 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.
test	The test to apply, that may be either "kpss" (Kwiatkowski, 1992) or "adf" (Dickey \& Fuller, 1981). If NULL (the default), the choice is for the KPSS test if the number of periods is less than 100, otherwise the ADF test is used.
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.
lag.order	The lag order at which the test statistic is computed. It is automatically selected according to the precedure by Ng \& Perron (2001).
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 ADF test is the presence of a unit root. The lag order to calculate the statistic of the ADF test is automatically selected according to the precedure by Ng \& Perron (2001).

The null hypothesis of the KPSS test is stationarity. The statistic of the KPSS test is calculated at the lag order 4*(n/100)^0.25.

If the group factor is specified, p-values of each group are combined using the method proposed by Demetrescu (2006).

References

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

D. Kwiatkowski, P. C. B. Phillips, P. Schmidt and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3): 159-178.

S. Ng, and W. P. Perron (2001). Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. *Econometrica*, 60: 1519-1554. DOI: 10.1111/1468-0262.00256.

Examples

```
data(industry)
indus.urt <- unirootTest(c("Job","Consum","Population","GDP"),
  group="Region",time="Year",data=industry,log=TRUE)
indus.urt      ## p-values
indus.urt$Job  ## details for variable 'Job'
```

Index