Package ‘codareg’

August 3, 2021

Type Package

Title Semi-Elasticities Computations for CoDa Regression Model

Version 1.0

Date 2021-07-31

Author Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan

Maintainer Thibault Laurent <Thibault.Laurent@tse-fr.eu>

Description Computes elasticities and semi-elasticities for compositional data models following the results derived on Morais and Thomas-Agnan (2020).

License GPL-2

LazyData true

Depends R (>= 3.5.0)

NeedsCompilation no

R topics documented:

`codareg-package` *Semi-Elasticities Computations for CoDa Regression Model*

Description

While impacts of covariates in standard regression models are straightforward and well known, this is not the case in compositional models. Therefor, this package computes elasticities and semi-elasticities for those compositional data models. Compositional variables can be included in the model both as endogenous and exogenous variables.

Some features of this package consist of: performance of the regression in the simplex and coordinate spaces through ALR and ILR transformations, estimation of impacts or (semi-)elasticities and its summarisation, and computation of p-values for the significance tests concerning regression parameters and (semi-)elasticities.

Thus, the main function of the package is `impacts`, whose input is the result of the regression obtained from `reg`.

It follows directly the results derived on Morais and Thomas-Agnan (2020).

Details

Index of help topics:

```
BDDSegX                 BDDSegX data.
F_D                     Computation of matrix F_D.
K_D                     Computation of matrix K_D.
P_D                     Computation of matrix P_D.
V_D                     Construction of the helmertian contrast
                        matrices.
alr                     ALR transformation.
```

alr_x_impacts	Semi-elasticities between compositional X and standard Y variables. ALR case.
alr_x_reg	Regression with compositional X and standard Y variables. ALR case.
alr_y_impacts	Semi-elasticities between standard X and compositional Y variables. ALR case.
alr_y_reg	Regression with standard X and compositional Y variables. ALR case.
alr_yx_impacts	Elasticities between compositional X and compositional Y variables. ALR case.
alr_yx_reg	Regression with compositional X and compositional Y variables. ALR case.
b_cov	Variance-covariance matrix of the coefficients.
check_formula	Translation of formula to suitable data.
closure	Closure operator.
codareg-package	Semi-Elasticities Computations for CoDa Regression Model
est_var	Estimation of the common covariance matrix.
ilr	ILR transformation.
ilr_x_impacts	Semi-elasticities between compositional X and standard Y variables. ILR case.
ilr_x_reg	Regression with compositional X and standard Y variables. ILR case.
ilr_y_impacts	Semi-elasticities between standard X and compositional Y variables. ILR case.
ilr_y_reg	Regression with standard X and compositional Y variables. ILR case.
ilr_yx_impacts	Elasticities between compositional X and compositional Y variables. ILR case.
ilr_yx_reg	Regression with compositional X and compositional Y variables. ILR case.
imp_summary	Summary of impacts.
impacts	Impacts of covariates for compositional models.
inner_reg	Internal regression function.
inv_permutation_D	Inverse of 'permutation_D'.
inversealr	Inverse of the ALR transformation.
inverseilr	Inverse of the ILR transformation.
mlm	Computes the estimated coefficients of the multivariate linear regression.
p_values_coef	P-values for coefficients via bootstrap.
p_values_se	P-values for (semi-)elasticities via bootstrap.
permutation_D	Move a column to the last position.
reg	Compositional regression function.
sd_se_x	S.d. for semi-elasticities with non-compositional Y.
share_ratio	Share ratio.
tidy	Tidy.
var2alr	Transform every variable to ALR coordinates.
var2ilr	Transform every variable to ILR coordinates.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan

Maintainer: Thibault Laurent <Thibault.Laurent@tse-fr.eu>

References

Morais, J., Thomas-Agnan, C., & Simioni, M. (2018). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, **47**(5), 1-25. doi: 10.17713/ajs.v47i5.718

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 139-160. ISBN: 940108324X.

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. ISBN 9780367453442.

Examples

```
# load provided data
data(BDDSegX)

# perform the regression
reg_results = reg(BDDSegX,
    formula = cbind(S_A,S_B,S_C,S_D,S_E)
    ~ PIB_Courant_t + HT_Gazole + PAC)

# obtain the impacts
imp_results = impacts(reg_results)

# display the summary of impacts
imp_summary(imp_results)
```

`alr` *ALR transformation.*

Description

Performs the ALR transformation to a matrix containing compositional data.

Usage

```
alr(x, R = dim(x)[2])
```

Arguments

`x`	Matrix containing the compositional data. If its rows do not add up to 1, the closure is applied.
`R`	Integer that indicates which part of `x` is desired to be taken as reference. If it is not specified, the last variable is used.

Value

A (D-1)x(D-1) matrix corresponding to the transformation of matrix `x` to the coordinate space through the ALR, where D is the number of compositional parts of `x`.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 139-160. ISBN: 940108324X.

See Also

See ilr for other transformation and inversealr for the inverse transformation.

Examples

```
compo_x = rbind(c(0.25,0.3,0.2),
                c(0.5,0.2,0.3),
                c(0.9,0.05,0.05),
                c(0.33,0.4,0.24),
                c(0.2,0.3,0.5))
alr(compo_x)
```

---

alr_x_impacts — *Semi-elasticities between compositional X and standard Y variables. ALR case.*

Description

Computes semielasticities in the model where the endogenous variable is standard and features -at least- compositional exogenous variables (may feature exogenous standard ones too). The additive logratio transformation (ALR) is carried out for the regression in the coordinate space.

Usage

```
alr_x_impacts(results)
```

Arguments

results	Named list containing the regression results which is returned by the `alr_x_reg()` function. It contains the matrix of coefficients in the coordinates, -which is required to compute these impacts-, among others.

Details

This is an 'internal' function whose aim is to be part of the general `impacts()` function. Therefore, the use of `impacts()` is recommended instead.

Value

A matrix of dimension DxP is returned where P is the number of components of the -possibly multivariate- endogenous variable and D stands for the sum of the number of parts of all the exogenous variables that are compositional.

Thus, element (i,j) corresponds to the computed semi-elasticity between the i-th part of the compositional variables and the j-th component of the endogenous variable.

Note

Functions `alr_x_impacts()` and `ilr_x_impacts()` obtain the same results for a given model, as these semi-elasticities are invariant to the chosen transformation. However, two different functions have been developed because their arguments are not the same, as they include the matrix of coefficients in the coordinate space.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Morais, J., Thomas-Agnan, C., & Simioni, M. (2018). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, **47**(5), 1-25. doi: 10.17713/ajs.v47i5.718

See Also

See `impacts` for the main function concerning (semi-)elasticities computation.

---

alr_x_reg — *Regression with compositional X and standard Y variables. ALR case.*

Description

Performs the regression of the model where the endogenous variable is standard and features -at least- compositional exogenous variables (may feature exogenous standard ones too). The additive logratio transformation (ALR) is carried out.

Usage

```
alr_x_reg(Y, X, Z = list(matrix()), R = list(NULL), constant = TRUE)
```

Arguments

`Y`	Data from the endogenous variable. It is a DxN matrix where D is the number of components of the -possibly- multivariate endogenous variable and N is the number of statistical units.
`X`	Data from exogenous variables which are compositional. It is a list of (P_k)xN matrices. The k-th matrix corresponds to the k-th compositional exogenous variable. P_k stands for the number of parts that features the k-th variable.
`Z`	Data from exogenous variables which are not compositional (therefore standard). It is a list of (Q_k)xN matrices. The k-th matrix corresponds to the k-th standard exogenous variable. Q_k stands for the number of components that features the k-th variable, which may be multivariate or not.
`R`	List of integers where the i-th element indicates which part of the i-th compostional variable is desired to be used as reference for the transformation. The order taken into account is the order of appearance of each compositional variable (from left to right) in the function. If `R` is not specified, the last part of each compositional variable is taken as reference.
`constant`	Logical variable. `TRUE` if an intercept is desired and `FALSE` otherwise.

Details

This is an 'internal' function whose aim is to be part of the general `reg()` function. Therefore, the use of `reg()` is recommended instead.

Value

Named list whose `"reg_type"` attribute is "alr_x_reg" and contains:

`Y_coord`	Endogenous data in the coordinate space.
`X_coord`	Exogenous data in the coordinate space.
`constant`	Logical that indicates if the model includes an intercept.
`B_coord`	Matrix of coefficients in the coordinate space.
`B_simplex`	Matrix of coefficients in the simplex space.
`Bcoord_cov`	Covariance matrix of coefficients in the coordinate space.
`fitted_v_coord`	Fitted values of the model in the coordinate space
`residuals_coord`	Residuals of the model in the coordinate space
`xdim`	List whose i-th element indicates the number of parts of the i-th exogenous compositional variable.

Note

Functions `alr_x_reg()` and `ilr_x_reg()` obtain the same results in the simplex for a given model, but not in the coordinates, as the transformation carried out is different.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. ISBN 9780367453442.

Nguyen, T. H. A., Laurent, T., Thomas-Agnan, C., & Ruiz-Gazen, A. (2020). Analyzing the impacts of socio-economic factors on French departmental elections with CoDa methods. *Journal of Applied Statistics*, 1-17. doi: 10.1080/02664763.2020.1858274

See Also

See `reg` for the main regression function.

alr_yx_impacts — *Elasticities between compositional X and compositional Y variables. ALR case.*

Description

Computes elasticities and semi-elasticities in the model where the endogenous variable is compositional and features -at least- compositional exogenous variables (may feature exogenous standard ones too). The additive logratio transformation (ALR) is carried out for the regression in the coordinate space.

Usage

```
alr_yx_impacts(results)
```

Arguments

`results`	Named list containing the regression results which is returned by the `alr_yx_reg()` function. It contains the matrix of coefficients in the coordinates, -which is required to compute these impacts-, among others.

Details

This is an 'internal' function whose aim is to be part of the general `impacts()` function. Therefore, the use of `impacts()` is recommended instead.

Value

A list of N matrices of dimension DxP where P is the number of parts of the endogenous variable, D stands for the total number of components of the exogenous variables (including both compositional and standard ones) and N is the number of observations. Thus, the i-th matrix corresponds to the i-th statistical unit, as these (semi-)elasticities depends on the value of the exogenous variables and therefore the matrix is unique for each observation.

Thus, element (i,j) corresponds to the computed (semi-)elasticity between the i-th component of the exogenous variables and the j-th part of the compositional endogenous variable.

Note

Functions `alr_yx_impacts()` and `ilr_yx_impacts()` obtain the same results for a given model, as these (semi-)elasticities are invariant to the chosen transformation. However, two different functions have been developed because their arguments are not the same, as they include the matrix of coefficients in the coordinate space.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Morais, J., Thomas-Agnan, C., & Simioni, M. (2018). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, **47**(5), 1-25. doi: 10.17713/ajs.v47i5.718

See Also

See `impacts` for the main function concerning (semi-)elasticities computation.

`alr_yx_reg` *Regression with compositional X and compositional Y variables. ALR case.*

Description

Performs the regression of the model where the endogenous variable is compositional and features -at least- compositional exogenous variables (may feature exogenous standard ones too). The additive logratio transformation (ALR) is carried out.

Usage

```
alr_yx_reg(Y, X, Z = list(matrix()), R = list(NULL), constant = TRUE)
```

Arguments

`Y`	Data from the endogenous variable. It is a DxN matrix where D is the number of parts of the compositional endogenous variable and N is the number of statistical units.
`X`	Data from exogenous variables which are compositional. It is a list of (P_k)xN matrices. The k-th matrix corresponds to the k-th compositional exogenous variable. P_k stands for the number of parts that features the k-th variable.
`Z`	Data from exogenous variables which are not compositional (therefore standard). It is a list of (Q_k)xN matrices. The k-th matrix corresponds to the k-th standard exogenous variable. Q_k stands for the number of components that features the k-th variable, which may be multivariate or not.
`R`	List of integers where the i-th element indicates which part of the i-th compositional variable is desired to be used as reference for the transformation. The order taken into account is the order of appearance of each compositional variable (from left to right) in the function. If `R` is not specified, the last part of each compositional variable is taken as reference.
`constant`	Logical variable. `TRUE` if an intercept is desired and `FALSE` otherwise.

Details

This is an 'internal' function whose aim is to be part of the general `reg()` function. Therefore, the use of `reg()` is recommended instead.

Value

Named list whose `"reg_type"` attribute is "alr_yx_reg" and contains:

`Y_coord`	Endogenous data in the coordinate space.
`X_coord`	Exogenous data in the coordinate space.
`constant`	Logical that indicates if the model includes an intercept.

`B_coord`	Matrix of coefficients in the coordinate space.
`B_simplex`	Matrix of coefficients in the simplex space.
`Bcoord_cov`	Covariance matrix of coefficients in the coordinate space.
`fitted_v_coord`	Fitted values of the model in the coordinate space
`residuals_coord`	Residuals of the model in the coordinate space

Note

Functions `alr_yx_reg()` and `ilr_yx_reg()` obtain the same results in the simplex for a given model, but not in the coordinates, as the transformation carried out is different.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. ISBN 9780367453442.

Nguyen, T. H. A., Laurent, T., Thomas-Agnan, C., & Ruiz-Gazen, A. (2020). Analyzing the impacts of socio-economic factors on French departmental elections with CoDa methods. *Journal of Applied Statistics*, 1-17. doi: 10.1080/02664763.2020.1858274

See Also

See `reg` for the main regression function.

---

alr_y_impacts — *Semi-elasticities between standard X and compositional Y variables. ALR case.*

Description

Computes semielasticities in the model where the endogenous variable is compositional and features classical exogenous variables (non-compositional). The additive logratio transformation (ALR) is carried out for the regression in the coordinate space.

Usage

```
alr_y_impacts(results)
```

Arguments

`results`	Named list containing the regression results which is returned by the `alr_y_reg()` function. It contains the matrix of coefficients in the coordinates, -which is required to compute these impacts-, among others.

Details

This is an 'internal' function whose aim is to be part of the general `impacts()` function. Therefore, the use of `impacts()` is recommended instead.

Value

A list of N matrices of dimension DxP where P is the number of parts of the endogenous variable, D stands for the total number of components of the exogenous variables and N the number of observations. Thus, the k-th matrix corresponds to the k-th statistical unit, as these semi-elasticities depends on the value of the exogenous variables and therefore the matrix is unique for each observation.

Therefore, element (i,j) of each matrix corresponds to the computed semi-elasticity between the i-th component of the exogenous variables and the j-th part of the endogenous one.

Note

Functions `alr_y_impacts()` and `ilr_y_impacts()` obtain the same results for a given model, as these semi-elasticities are invariant to the chosen transformation. However, two different functions have been developed because their arguments are not the same, as they include the matrix of coefficients in the coordinate space.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Morais, J., Thomas-Agnan, C., & Simioni, M. (2018). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, **47**(5), 1-25. doi: 10.17713/ajs.v47i5.718

See Also

See `impacts` for the main function concerning (semi-)elasticities computation.

`alr_y_reg` *Regression with standard X and compositional Y variables. ALR case.*

Description

Performs the regression of the model where the endogenous variable is compositional and features standard exogenous variables. The additive logratio transformation (ALR) is carried out.

Usage

```
alr_y_reg(Y, X, R = list(NULL), constant = TRUE)
```

Arguments

Y	Data from the endogenous variable. It is a DxN matrix where D is the number of parts of the endogenous variable and N is the number of statistical units.
X	Data from exogenous variables which are standard (non-compositional). It is a list of (P_k)xN matrices. The k-th matrix corresponds to the k-th exogenous variable. P_k stands for the number of components that features the k-th variable (that may be multivariate).

R	List of one integer that indicates which part of the endogenous compostional variable is desired to be used as reference for the transformation. If R is not specified, the last part of the variable is used.
constant	Logical variable. TRUE if an intercept is desired and FALSE otherwise.

Details

This is an 'internal' function whose aim is to be part of the general reg() function. Therefore, the use of reg() is recommended instead.

Value

Named list whose "reg_type" attribute is "alr_y_reg" and contains:

Y_coord	Endogenous data in the coordinate space.
X_coord	Exogenous data in the coordinate space.
constant	Logical that indicates if the model includes an intercept.
B_coord	Matrix of coefficients in the coordinate space.
B_simplex	Matrix of coefficients in the simplex space.
Bcoord_cov	Covariance matrix of coefficients in the coordinate space.
fitted_v_coord	Fitted values of the model in the coordinate space
residuals_coord	Residuals of the model in the coordinate space

Note

Functions alr_y_reg() and ilr_y_reg() obtain the same results in the simplex for a given model, but not in the coordinates, as the transformation carried out is different.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. ISBN 9780367453442.

Nguyen, T. H. A., Laurent, T., Thomas-Agnan, C., & Ruiz-Gazen, A. (2020). Analyzing the impacts of socio-economic factors on French departmental elections with CoDa methods. *Journal of Applied Statistics*, 1-17. doi: 10.1080/02664763.2020.1858274

See Also

See reg for the main regression function.

BDDSegX *BDDSegX data.*

Description

Example dataset about the automotive market in .rdata format.

Usage

```
BDDSegX
```

Format

A data frame with 152 observation on 21 variables.

The main variables, which correspond to different market shares of the automotive market -and that are named `S_A`, `S_B`, `S_C`, `S_D`, and `S_E`- were simulated from private data. Below there is a description of the variables:

- `S_A`: A segment market share (in terms of sales volume). Period: Month. Source: Renault (simulated).
- `S_B`: B segment market share (in terms of sales volume). Period: Month. Source: Renault (simulated).
- `S_C`: C segment market share (in terms of sales volume). Period: Month. Source: Renault (simulated).
- `S_D`: D segment market share (in terms of sales volume). Period: Month. Source: Renault (simulated).
- `S_E`: E segment market share (in terms of sales volume). Period: Month. Source: Renault (simulated).
- `DC_Men_Courant`: Household consumption expenditure - Total - Volume at previous year's prices (chained) - Serie CVS-CJO (Millions euro). Period: Quarter. Source: INSEE.
- `Eco_Sentim_Indic`: Made up of the 15 individual components of the confidence indicators for Construction (5%), Services (30%), Retail trade (5%), Industry (40%) and Consumers (20%). Weights are applied to the standardised individual component series of the five confidence indicators. Period: Month. Source: EUROSTAT.
- `FBCF_Men_Courant`: GFCF of households - Total - Volume at previous year's prices (chained) - Serie CVS-CJO (Millions euro). Period: Quarter. Source: INSEE.
- `PAC`: Scrapping incentive (dummy variable). Period: Month. Source: Wikipedia.
- `PIB_Courant_t`: Total GDP at current prices - Serie CVS-CJO (Millions euro). Period: Quarter. Source: INSEE.
- `TTC_Gazole`: Mean TTC price of diesel (France, euro per litre). Period: Month. Source: MEDDE.
- `HC_Gazole`: Mean HT price of diesel (France, euro per litre). Period: Month. Source: MEDDE.
- `Tx_int_CT`: Short-term interest rates are the short-term borrowing rates applied between financial institutions or the rates on short-term government securities on the primary or secondary market. Short-term interest rates are usually averages of daily rates, expressed as a percentage. Short-term interest rates are three-month money market rates, when available. They are commonly referred to by the standard terms "taux du marche monetaire" and "taux des bons du Tresor". (%, France). Period: Month. Source: OCDE.

- `In_Share_Lowest`: Share of income that belongs to housholds in percentiles 0-20 of income distribution (%). Period: Annual. Source: Eurostat.
- `In_Share_2nd`: Share of income that belongs to housholds in percentiles 20-40 of income distribution (%). Period: Annual. Source: Eurostat.
- `In_Share_3rd`: Share of income that belongs to housholds in percentiles 40-60 of income distribution (%). Period: Annual. Source: Eurostat.
- `In_Share_4th`: Share of income that belongs to housholds in percentiles 60-80 of income distribution (%). Period: Annual. Source: Eurostat.
- `In_Share_high`: Share of income that belongs to housholds in percentiles 80-100 of income distribution (%). Period: Annual. Source: Eurostat.

References

Morais, J. and Thomas-Agnan, C. (2019), Impact of economic context on automobile market segment shares: a compositional approach, *Case Studies in Business, Industry and Government Statistics*, in press.

Examples

```
data(BDDSegX)
```

`b_cov` *Variance-covariance matrix of the coefficients.*

Description

Compute the variance-covariance matrix of the coefficients for any multivariate regression model.

Usage

```
b_cov(X, Y, B)
```

Arguments

`X`	Matrix with the exogenous data where columns are variables and rows are statistical units.
`Y`	Matrix with the endogenous data where columns are variables and rows are statistical units.
`B`	Matrix of estimated coefficients of the regression. If it is not specified, `mlm(Y,X)` is used.

Details

This is an 'internal' function whose aim is to be part of the general `reg()` function. Therefore, the use of `reg()`, whose output includes the coefficient variance-covariance matrix, is recommended.

Value

Returns the usual symmetric variance-covariance matrix where the element (i,j) corresponds to the covariance between coefficient "i" estimation and coefficient "j" estimation. In the multivariate case, denoting elements of B as by b_ij, columns and rows of the variance matrix are listed as (b_11, b_21, ..., b_P1, b_12, b_22, ..., b_P2, ... , b_PP) where P is the number of components of the endogenous variable.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. ISBN 9780367453442.

See Also

See `reg` for the main regression function and obtaining this matrix in practice.

`check_formula` *Translation of formula to suitable data.*

Description

This function takes the provided formula and dataset and converts variables indicated to lists of data that internal functions can handle.

It also checks that characteristics of the selected variables from the dataset are suitable to estimate that specific model. In particular, it checks that every element in a variable is a number and, in the compositional case, if that number is strictly positive.

Usage

```
check_formula(dataset, formula)
```

Arguments

`dataset`	Dataset in data.frame format.
`formula`	Formula of the desired model. Compositional variables should be inserted as a concatenation of their parts thorugh 'cbind()' and standard ones with a plus sign '+' between all of them, including multivariate ones.

Details

This is an 'internal' function whose aim is to be part of the general `reg()` function. Therefore there is no need to use this function separately from `reg()`.

Value

Returns a list which is used in the `reg()` function. This `check_formula()` function is an intermediate function.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

`closure` *Closure operator.*

Description

Compute the closure for compositional data.

Usage

```
closure(x, k)
```

Arguments

`x`	Matrix with compositional data. Each row represents one observation and each column, one part of the compositional variable.
`k`	Desired total of the closure. Each row sums up to 'k' once the closure is performed. If it is not specified, `k=1`.

Details

This function is inserted in different functions. Thus, it is possible to specify compositional variables whose total is not constant in all of them and the corresponding function will apply the closure automatically. Therefore, in general terms, the use of `closure()` is not necessary.

Value

Matrix with the transformed data, which belongs to the simplex space.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 139-160. ISBN: 940108324X.

`est_var` *Estimation of the common covariance matrix.*

Description

It is estimated the common covariance matrix variance following equation (5.24) in Kim & Timm (2006).

Usage

```
est_var(X, Y, B)
```

Arguments

`X`	Matrix with the exogenous variables of the regression.
`Y`	Matrix with the endogenous variable of the regression.
`B`	Matrix of coefficients of the regression. If missing, it is used B=mlm(Y,X).

Details

This function is used as an intermediate function in order to obtain the variance-covariance matrix of the coefficients.

Value

It is returned a DxD matrix (D is the number of components of the endogenous variable), where element (i,j) denotes the estimated covariance between components i and j.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. ISBN 9780367453442.

See Also

See `b_cov` for computing the variance-covariance matrix of the coefficients.

`F_D` *Computation of matrix F_D.*

Description

This function creates the (D-1)xD matrix 'F_D' that is needed to carry out some computations with respect to the ALR transformation. In particular, alr(x) = F_D(D) ln(x)' where D is the number of compositional parts of x.

Usage

```
F_D(D)
```

Arguments

`D`	Number of rows of the matrix.

Value

It is returned the matrix F_D = (Id | -1) where Id is the identity of dimension D-1 and -1 denotes the column vector of ones of dimension (D-1).

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

`ilr` *ILR transformation.*

Description

Performs the ILR transformation to a matrix containing compositional data.

Usage

```
ilr(x, V)
```

Arguments

`x`	Matrix containing the compositional data. If its rows do not add up to 1, the closure is applied.
`V`	Dx(D-1) contrast matrix, where D is the number of compositional parts of `x`. If it is not specified, `V_D(D)` is used.

Value

A (D-1)x(D-1) matrix corresponding to the transformation of matrix `x` to the coordinate space through the ILR.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 139-160. ISBN: 940108324X.

See Also

See `alr` for other transformation and `inverseilr` for the inverse transformation.

Examples

```
compo_x = rbind(c(0.25,0.3,0.2),
                c(0.5,0.2,0.3),
                c(0.9,0.05,0.05),
                c(0.33,0.4,0.24),
                c(0.2,0.3,0.5))
ilr(compo_x)
```

`ilr_x_impacts` *Semi-elasticities between compositional X and standard Y variables. ILR case.*

Description

Computes semielasticities in the model where the endogenous variable is standard and features -at least- compositional exogenous variables (may feature exogenous standard ones too). The isometric logratio transformation (ILR) is carried out for the regression in the coordinate space.

Usage

```
ilr_x_impacts(results)
```

Arguments

`results`	Named list containing the regression results which is returned by the `ilr_x_reg()` function. It contains the matrix of coefficients in the coordinates, -which is required to compute these impacts-, among others.

Details

This is an 'internal' function whose aim is to be part of the general `impacts()` function. Therefore, the use of `impacts()` is recommended instead.

Value

A matrix of dimension DxP is returned where P is the number of components of the -possibly multivariate- endogenous variable and D stands for the sum of the number of parts of all the exogenous variables that are compositional.

Thus, element (i,j) corresponds to the computed semi-elasticity between the i-th part of the compositional variables and the j-th component of the endogenous variable.

Note

Functions `alr_x_impacts()` and `ilr_x_impacts()` obtain the same results for a given model, as these semi-elasticities are invariant to the chosen transformation. However, two different functions have been developed because their arguments are not the same, as they include the matrix of coefficients in the coordinate space.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Morais, J., Thomas-Agnan, C., & Simioni, M. (2018). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, **47**(5), 1-25. doi: 10.17713/ajs.v47i5.718

See Also

See `impacts` for the main function concerning (semi-)elasticities computation.

`ilr_x_reg` *Regression with compositional X and standard Y variables. ILR case.*

Description

Performs the regression of the model where the endogenous variable is standard and features -at least- compositional exogenous variables (may feature exogenous standard ones too). The isometric logratio transformation (ILR) is carried out.

Usage

```
ilr_x_reg(Y, X, Z = list(matrix()), V = list(matrix()), constant = TRUE)
```

Arguments

`Y`	Data from the endogenous variable. It is a DxN matrix where D is the number of components of the -possibly- multivariate endogenous variable and N is the number of statistical units.
`X`	Data from exogenous variables which are compositional. It is a list of (P_k)xN matrices. The k-th matrix corresponds to the k-th compositional exogenous variable. P_k stands for the number of parts that features the k-th variable.
`Z`	Data from exogenous variables which are not compositional (therefore standard). It is a list of (Q_k)xN matrices. The k-th matrix corresponds to the k-th standard exogenous variable. Q_k stands for the number of components that features the k-th variable, which may be multivariate or not.
`V`	List where the i-th element indicates which contrast matrix is desired to be used as reference for the transformation. The order taken into account is the order of appearance of each compositional variable (from left to right) in the function. If `V` is not specified, `V_D(Q_k)` is used for the k-th variable.
`constant`	Logical variable. `TRUE` if an intercept is desired and `FALSE` otherwise.

Details

This is an 'internal' function whose aim is to be part of the general `reg()` function. Therefore, the use of `reg()` is recommended instead.

Value

Named list whose `"reg_type"` attribute is "ilr_x_reg" and contains:

`Y_coord` Endogenous data in the coordinate space.

`X_coord` Exogenous data in the coordinate space.

`constant` Logical that indicates if the model includes an intercept.

`B_coord` Matrix of coefficients in the coordinate space.

`B_simplex` Matrix of coefficients in the simplex space.

`Bcoord_cov` Covariance matrix of coefficients in the coordinate space.

`fitted_v_coord` Fitted values of the model in the coordinate space

`residuals_coord` Residuals of the model in the coordinate space

`xdim` List whose i-th element indicates the number of parts of the i-th exogenous compositional variable.

Note

Functions `alr_x_reg()` and `ilr_x_reg()` obtain the same results in the simplex for a given model, but not in the coordinates, as the transformation carried out is different.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. ISBN 9780367453442.

Nguyen, T. H. A., Laurent, T., Thomas-Agnan, C., & Ruiz-Gazen, A. (2020). Analyzing the impacts of socio-economic factors on French departmental elections with CoDa methods. *Journal of Applied Statistics*, 1-17. doi: 10.1080/02664763.2020.1858274

See Also

See `reg` for the main regression function.

`ilr_yx_impacts` *Elasticities between compositional X and compositional Y variables. ILR case.*

Description

Computes elasticities and semi-elasticities in the model where the endogenous variable is compositional and features -at least- compositional exogenous variables (may feature exogenous standard ones too). The isometric logratio transformation (ILR) is carried out for the regression in the coordinate space.

Usage

```
ilr_yx_impacts(results)
```

Arguments

`results`	Named list containing the regression results which is returned by the `ilr_yx_reg()` function. It contains the matrix of coefficients in the coordinates, -which is required to compute these impacts-, among others.

Details

This is an 'internal' function whose aim is to be part of the general `impacts()` function. Therefore, the use of `impacts()` is recommended instead.

Value

A list of N matrices of dimension DxP where P is the number of parts of the endogenous variable, D stands for the total number of components of the exogenous variables (including both compositional and standard ones) and N is the number of observations. Thus, the i-th matrix corresponds to the i-th statistical unit, as these (semi-)elasticities depends on the value of the exogenous variables and therefore the matrix is unique for each observation.

Thus, element (i,j) corresponds to the computed (semi-)elasticity between the i-th component of the exogenous variables and the j-th part of the compositional endogenous variable.

Note

Functions `alr_yx_impacts()` and `ilr_yx_impacts()` obtain the same results for a given model, as these (semi-)elasticities are invariant to the chosen transformation. However, two different functions have been developed because their arguments are not the same, as they include the matrix of coefficients in the coordinate space.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Morais, J., Thomas-Agnan, C., & Simioni, M. (2018). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, **47**(5), 1-25. doi: 10.17713/ajs.v47i5.718

See Also

See `impacts` for the main function concerning (semi-)elasticities computation.

`ilr_yx_reg` *Regression with compositional X and compositional Y variables. ILR case.*

Description

Performs the regression of the model where the endogenous variable is compositional and features -at least- compositional exogenous variables (may feature exogenous standard ones too). The isometric logratio transformation (ILR) is carried out.

Usage

```
ilr_yx_reg(Y, X, Z = list(matrix()), V = list(matrix()), constant = TRUE)
```

Arguments

`Y`	Data from the endogenous variable. It is a DxN matrix where D is the number of parts of the compositional endogenous variable and N is the number of statistical units.
`X`	Data from exogenous variables which are compositional. It is a list of (P_k)xN matrices. The k-th matrix corresponds to the k-th compositional exogenous variable. P_k stands for the number of parts that features the k-th variable.
`Z`	Data from exogenous variables which are not compositional (therefore standard). It is a list of (Q_k)xN matrices. The k-th matrix corresponds to the k-th standard exogenous variable. Q_k stands for the number of components that features the k-th variable, which may be multivariate or not.
`V`	List where the i-th element indicates which contrast matrix is desired to be used as reference for the transformation. The order taken into account is the order of appearance of each compositional variable (from left to right) in the function. If `V` is not specified, `V_D(D)` is used for the endogenous variable and `V_D(Q_k)` is used for the k-th compositional exogenous variable.
`constant`	Logical variable. `TRUE` if an intercept is desired and `FALSE` otherwise.

Details

This is an 'internal' function whose aim is to be part of the general `reg()` function. Therefore, the use of `reg()` is recommended instead.

Value

Named list whose `"reg_type"` attribute is "ilr_yx_reg" and contains:

`Y_coord`	Endogenous data in the coordinate space.
`X_coord`	Exogenous data in the coordinate space.
`constant`	Logical that indicates if the model includes an intercept.
`B_coord`	Matrix of coefficients in the coordinate space.

`B_simplex`	Matrix of coefficients in the simplex space.
`Bcoord_cov`	Covariance matrix of coefficients in the coordinate space.
`fitted_v_coord`	Fitted values of the model in the coordinate space
`residuals_coord`	Residuals of the model in the coordinate space

Note

Functions `alr_yx_reg()` and `ilr_yx_reg()` obtain the same results in the simplex for a given model, but not in the coordinates, as the transformation carried out is different.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. ISBN 9780367453442.

Nguyen, T. H. A., Laurent, T., Thomas-Agnan, C., & Ruiz-Gazen, A. (2020). Analyzing the impacts of socio-economic factors on French departmental elections with CoDa methods. *Journal of Applied Statistics*, 1-17. doi: 10.1080/02664763.2020.1858274

See Also

See `reg` for the main regression function.

`ilr_y_impacts` *Semi-elasticities between standard X and compositional Y variables. ILR case.*

Description

Computes semielasticities in the model where the endogenous variable is compositional and features classical exogenous variables (non-compositional). The isometric logratio transformation (ILR) is carried out for the regression in the coordinate space.

Usage

```
ilr_y_impacts(results)
```

Arguments

`results`	Named list containing the regression results which is returned by the `ilr_y_reg()` function. It contains the matrix of coefficients in the coordinates, -which is required to compute these impacts-, among others.

Details

This is an 'internal' function whose aim is to be part of the general `impacts()` function. Therefore, the use of `impacts()` is recommended instead.

Value

A list of N matrices of dimension DxP where P is the number of parts of the endogenous variable, D stands for the total number of components of the exogenous variables and N the number of observations. Thus, the i-th matrix corresponds to the i-th statistical unit, as these semi-elasticities depends on the value of the exogenous variables and therefore the matrix is unique for each observation.

Therefore, element (i,j) of each matrix corresponds to the computed semi-elasticity between the i-th component of the exogenous variables and the j-th part of the endogenous one.

Note

Functions `alr_y_impacts()` and `ilr_y_impacts()` obtain the same results for a given model, as these semi-elasticities are invariant to the chosen transformation. However, two different functions have been developed because their arguments are not the same, as they include the matrix of coefficients in the coordinate space.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Morais, J., Thomas-Agnan, C., & Simioni, M. (2018). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, **47**(5), 1-25. doi: 10.17713/ajs.v47i5.718

See Also

See `impacts` for the main function concerning (semi-)elasticities computation.

`ilr_y_reg` *Regression with standard X and compositional Y variables. ILR case.*

Description

Performs the regression of the model where the endogenous variable is compositional and features standard exogenous variables. The isometric logratio transformation (ILR) is carried out.

Usage

```
ilr_y_reg(Y, X, V = list(matrix()), constant = TRUE)
```

Arguments

`Y`	Data from the endogenous variable. It is a DxN matrix where D is the number of parts of the endogenous variable and N is the number of statistical units.
`X`	Data from exogenous variables which are standard (non-compositional). It is a list of (P_k)xN matrices. The k-th matrix corresponds to the k-th exogenous variable. P_k stands for the number of components that features the k-th variable (that may be multivariate).
`V`	List of one matrix that indicates which contrast matrix is desired to be used for the transformation. If it is not specified, `V_D(D)` is used.
`constant`	Logical variable. `TRUE` if an intercept is desired and `FALSE` otherwise.

Details

This is an 'internal' function whose aim is to be part of the general `reg()` function. Therefore, the use of `reg()` is recommended instead.

Value

Named list whose `"reg_type"` attribute is "ilr_y_reg" and contains:

`Y_coord`	Endogenous data in the coordinate space.
`X_coord`	Exogenous data in the coordinate space.
`constant`	Logical that indicates if the model includes an intercept.
`B_coord`	Matrix of coefficients in the coordinate space.
`B_simplex`	Matrix of coefficients in the simplex space.
`Bcoord_cov`	Covariance matrix of coefficients in the coordinate space.
`fitted_v_coord`	Fitted values of the model in the coordinate space
`residuals_coord`	Residuals of the model in the coordinate space

Note

Functions `alr_y_reg()` and `ilr_y_reg()` obtain the same results in the simplex for a given model, but not in the coordinates, as the transformation carried out is different.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. ISBN 9780367453442.

Nguyen, T. H. A., Laurent, T., Thomas-Agnan, C., & Ruiz-Gazen, A. (2020). Analyzing the impacts of socio-economic factors on French departmental elections with CoDa methods. *Journal of Applied Statistics*, 1-17. doi: 10.1080/02664763.2020.1858274

See Also

See `reg` for the main regression function.

`impacts` *Impacts of covariates for compositional models.*

Description

This is the main function of the package. It computes impacts of covariates for any compositional model through transformations ALR and ILR (although computations are different, results are invariant with respect to the transformation). Internally, this function recognizes which type of variables are included in the given model and calls the appropiate function: `ilr_x_impacts()`, `ilr_y_impacts()`, `alr_x_impacts()`, etc.

Usage

```
impacts(results, p_values = FALSE)
```

Arguments

`results`	Named list containing the regression results which is returned by the `reg()` function.
`p_values`	In the case that compositional variables included are only explanatory, it is possible to obtain the variance of the semi-elasticities and therefore to compute a t-test. Thus, this is a logical variable and, if `p-values = TRUE`, the usual significance stars is displayed next to each semi-elasticity.

Details

Results on impacts do not depend on the transformation chosen or their reference or contrast matrices.

Value

It depends on the variables included, as explained in the reference. Mainly, if the endogenous variable is compositional, (semi-)elasticities are different for each statistical unit.

Thus, if there are some exogenous compositional variables and the endogenous one is also compositional, see `alr_xy_impacts()` or `ilr_xy_impacts()` functions. If there are some exogenous compositional variables, but the endogenous one is standard, see `alr_x_impacts()` or `ilr_x_impacts()`. Finally, if the unique compositional variable is the endogenus one, see `alr_y_impacts()` or `ilr_y_impacts()`.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Morais, J., Thomas-Agnan, C., & Simioni, M. (2018). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, **47**(5), 1-25. doi: 10.17713/ajs.v47i5.718

See Also

For a summary of these (semi-)elasticities, see `imp_summary`. For the individual computation of p-values via bootstrapping in cases where the endogenous variable is compositional, see `p_values_se`. For the computation of p-values via t-test when the exogenous is not compositional see `sd_se_x`.

Examples

```
# Load data
data(BDDSegX)

# Obtain regression results
reg_results = reg(BDDSegX, formula =
      cbind(S_A,S_B,S_C,S_D,S_E) ~ PIB_Courant_t + PAC)

# Obtain impacts' result. In this example only
# the first 4 semi-elasticities are displayed.
imp_results = impacts(reg_results)
```

```
imp_results[1:4]
```

---

imp_summary *Summary of impacts.*

Description

This function creates a summary of impacts for models where the endogenous variable is compositional. Impacts for those cases depend on values of the explanatory variables, and therefore each observation features different impacts. Thus, maximums, minimums, and 1st, 2nd and 3rd quantiles of each impact's distribution are displayed.

Usage

```
imp_summary(impact_results)
```

Arguments

impact_results	Result of the function `impacts()`.

Value

Matrix where mean, maximum, minimum, and the usual quantiles of the distribution of (semi-)elasticities between the exogenous and endogenous variables are shown as long as one of them is compositional.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

Examples

```
# load data
data(BDDSegX)

# obtain regression results
reg_results    = reg(BDDSegX, formula =
      cbind(S_A,S_B,S_C,S_D,S_E) ~ PIB_Courant_t + PAC)

# obtain impacts results
impact_results = impacts(reg_results)

# display the summary of impacts
imp_summary(impact_results)
```

inner_reg *Internal regression function.*

Description

This function performs the estimation of coefficients in models where compositional variables are involved. Its results are used as an input for the computation of p-values via bootstrap in the main regression function `reg()`.

Usage

```
inner_reg(dataset, formula, transformation = "ILR", V = list(matrix()), R = list(NULL))
```

Arguments

`dataset`	Dataset in data.frame format.
`formula`	Formula of the desired model. Compositional variables should be inserted as a concatenation of their parts thorugh 'cbind()' and standard ones, although multivariate, with a plus '+' sign between all of them.
`transformation`	This argument can be "ILR" or "ALR" and indicates which type of transformation is employed.
`V`	This argument is only used if `transformation="ILR"`. It is a list. Each element of V is a matrix that is desired to be employed in the ILR transformation. The variable for which the i-th matrix is used is the i-th compositional variable, in order of appearance in the formula. If no list V is determined, contrast matrices constructed using the Helmert transformation are used.
`R`	This argument is only used if `transformation="ALR"`. It is a list of integers. Each element of R designs which part of the compositional variable is desired to be taken in the ALR transformation as reference. The variable for which the i-th integer is used is the i-th compositional variable, in order of appearance in the formula. If it is not specified, the last part of each variable is taken as reference.

Details

As mentioned, this is an 'intermediate' function and therefore it is not intended to be used directly.

Value

It depends on the nature of the variables included.

Thus, if there are some exogenous compositional variables and the endogenous one is also compositional, see `alr_xy_reg()` or `ilr_xy_reg()` functions, depending on the case. If there are some exogenous compositional variables, but the endogenous one is standard, see `alr_x_reg()` or `ilr_x_reg()`. Finally, if the unique compositional variable is the endogenus one, see `alr_y_reg()` or `ilr_y_reg()`.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

See Also

For the main regression function, see `reg`.

`inversealr` *Inverse of the ALR transformation.*

Description

Performs the ALR inverse transformation to a matrix containing data in the coordinate space.

Usage

```
inversealr(x, R = dim(x)[2] + 1)
```

Arguments

`x`	Matrix data of dimenstion (D-1)xN where D is the number of parts of the original data in the simplex and N the number of statistical units.
`R`	Integer that indicates which part of the original `x` in the simplex is taken as reference. If it is not specified, the last variable is used.

Value

A DxD matrix corresponding to the transformation of matrix `x` to the simplex space through the inverse of the ALR.

Note

Identity `inversealr(alr(x))==x` holds if and only if `x` is a matrix containing compositional data.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 139-160. ISBN: 940108324X.

See Also

See `alr` for the original transformation.

`inverseilr` *Inverse of the ILR transformation.*

Description

Performs the ILR inverse transformation to a matrix containing data in the coordinate space.

Usage

```
inverseilr(x, V)
```

Arguments

`x`	Matrix data of dimenstion (D-1)xN where D is the number of parts of the original data in the simplex and N the number of statistical units.
`V`	Dx(D-1) contrast matrix. If it is not specified, `V_D(D)` is used.

Value

A DxD matrix corresponding to the transformation of matrix `x` to the simplex space through the inverse of the ILR.

Note

Identity `inverseilr(ilr(x))==x` holds if and only if `x` is a matrix containing compositional data.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 139-160. ISBN: 940108324X.

See Also

See `ilr` for the original transformation.

inv_permutation_D *Inverse of 'permutation_D'.*

Description

It is an intermediate function which undoes the permutation made by permutation_D.

Usage

```
inv_permutation_D(mat, P)
```

Arguments

mat	Data matrix with the variable to be permuted in the last column.
P	Position that the last column will occupy in the output matrix.

Value

A new matrix with the last column of 'mat' in position P. Variables that were in positions P, P+1, P+2... occupy in this matrix positions P+1, P+2, P+3...

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

See Also

For the original permutation see permutation_D.

K_D *Computation of matrix K_D.*

Description

This function creates the Dx(D-1) matrix 'K_D' that is needed to carry out some computations with respect to the ALR transformation.

Usage

```
K_D(D)
```

Arguments

D	Number of rows of the matrix.

Value

It is returned the matrix K_D which is defined as such in the reference.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Chakir, R., Laurent, T., Ruiz-Gazen, A., Thomas-Agnan, C. & Vignes, C. (2017). Land Use Predictions on a Regular Grid at Different Scales and with Easily Accessible Covariates. *Revue economique*, **68**, 435-469. doi: 10.3917/reco.683.0435

`mlm` *Computes the estimated coefficients of the multivariate linear regression.*

Description

This function computes the BLUE estimator of the coefficients of the parameters for the multivariate general model as B_hat = (X'X)^(-1)X'Y as shown in the reference at equation (5.13).

Usage

```
mlm(Y, X)
```

Arguments

`Y`	It is a matrix that contains the endogenous variable data, which could be multivariate. Each row is an observation and each column a component of the variable.
`X`	It is a matrix that contains the exogenous variables data, which could be multivariate. Each row is an observation and each column a component of the variable.

Details

If Y is univariate, this model corresponds to the usual multiple linear regression one.

This function is integrated in many others -specially, in `reg()`, and it is not intended for direct use.

Value

The function returns the matrix B_hat of estimated coefficients of the parameters.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. ISBN 9780367453442.

See Also

For the estimation of the variance matrix see `b_cov`, and for the regression concerning compositional models, `reg`.

permutation_D *Move a column to the last position.*

Description

In order to compute the ALR transformation for a given reference variable other than the last one a couple of permutations should be carried out. This permutation takes the column in a certain position of the given matrix and returns the same matrix but with the specified column in the last position in order to be used as the reference for the ALR transformation.

Usage

```
permutation_D(mat, P)
```

Arguments

`mat`	Data matrix.
`P`	Position that occupies the column to be permuted to be the moved to the last column of the matrix.

Value

A new matrix where the column of 'mat' in position P is returned as the last column. Variables that were in positions P+1, P+2, P+3... occupy in this matrix positions P, P+1, P+2...

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

See Also

For the inverse permutation, see `inv_permutation_D`.

P_D *Computation of matrix P_D.*

Description

This Dx(D-1) matrix 'P_D' is needed to carry out some computations with respect the inverse ALR transformation.

Usage

```
P_D(D)
```

Arguments

`D`	Number of rows of the matrix, which is Dx(D-1).

Value

It is returned the matrix P_D which is the identity of dimension D-1 plus a last row of zeros which is a row vector 1x(D-1).

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

`p_values_coef` *P-values for coefficients via bootstrap.*

Description

This function computes the p-values associated to the test of significance of the coefficients via bootstrapping for any model. This is how it works: `B` samples with replacement are taken from the original dataset and the coefficients associated to the sample are computed. The p-value is twice the maximum between the proportion of coefficients that are larger than some threshold and 1 minus that proportion.

The threshold depends on the nature of the variables that are associated to the coefficient. If the coefficient relates a compositional exogenous variable to a compositional endogenous one, this threshold is 0. If it relates a compositional exogenous variable and a standard endogenous one, it is 1/P where P is the number of parts of the exogenous. If, on contrary, the coefficient is associated to a exogenous variable that is standard and the endogenous is compositional, the threshold is 1/D, where D is the number of parts of the endogenous.

Usage

```
p_values_coef(dataset, formula, transformation = "ILR",
      V = list(matrix()), R = list(NULL), B = 1000, seed = 31000)
```

Arguments

`dataset`	Dataset in data.frame format.
`formula`	Formula of the desired model. Compositional variables should be inserted as a concatenation of their parts thorugh 'cbind()' and standard ones, although multivariate, with a plus '+' sign between all of them.
`transformation`	This argument can be "ILR" or "ALR" and indicates which type of transformation is employed.
`V`	This argument is only used if `transformation="ILR"`. It is a list. Each element of V is a matrix that is desired to be employed in the ILR transformation. The variable for which the i-th matrix is used is the i-th compositional variable, in order of appearance in the formula. If no list V is determined, contrast matrices constructed using the Helmert transformation are used.
`R`	This argument is only used if `transformation="ALR"`. It is a list of integers. Each element of R designs which part of the compositional variable is desired to be taken in the ALR transformation as reference. The variable for which the i-th integer is used is the i-th compositional variable, in order of appearance in the formula. If it is not specified, the last part of each variable is taken as reference.
`B`	It indicates the number of bootstrap samples.
`seed`	It is a logical variable that only works if `pres = TRUE` and `simplex_pvalues = "bootstrap"`. It indicates the seed for the sampling.

Details

Although this function is designed for integration within `reg()`, its direct use can be useful, if the exact value of the p-values is desired.

Value

A matrix of the same dimension and layout as that of coefficients is returned. Each element (i,j) of this p-values matrix shows the p-value associated to the element (i,j) in the matrix of coefficients.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Nguyen, T. H. A., Laurent, T., Thomas-Agnan, C., & Ruiz-Gazen, A. (2020). Analyzing the impacts of socio-economic factors on French departmental elections with CoDa methods. *Journal of Applied Statistics*. Supplementary material. doi: 10.1080/02664763.2020.1858274

Examples

```
# Load data
data(BDDSegX)

# Obtain the matrix of p-values
p_values_coef(BDDSegX, formula =
      cbind(S_A,S_B,S_C,S_D,S_E) ~ PIB_Courant_t + PAC)
```

`p_values_se` *P-values for (semi-)elasticities via bootstrap.*

Description

This function computes the p-values associated to the significance test of the (semi-)elasticities in models where the endogenous is compositional and for a given statistical unit. This is done through bootstrapping.

Usage

```
p_values_se(dataset, formula, observation, transformation = "ILR",
    V = list(matrix()), R = list(NULL), B = 1000, seed = 31000)
```

Arguments

`dataset`	Dataset in data.frame format.
`formula`	Formula of the desired model. Compositional variables should be inserted as a concatenation of their parts thorugh 'cbind()' and standard ones, although multivariate, with a plus '+' sign between all of them.
`observation`	It is an integer that indicates the index number of that statistical unit for which it is desired to compute the p-values.

`transformation` This argument can be "ILR" or "ALR" and indicates which type of transformation is employed.

`V` This argument is only used if `transformation="ILR"`. It is a list. Each element of V is a matrix that is desired to be employed in the ILR transformation. The variable for which the i-th matrix is used is the i-th compositional variable, in order of appearance in the formula. If no list V is determined, contrast matrices constructed using the Helmert transformation are used.

`R` This argument is only used if `transformation="ALR"`. It is a list of integers. Each element of R designs which part of the compositional variable is desired to be taken in the ALR transformation as reference. The variable for which the i-th integer is used is the i-th compositional variable, in order of appearance in the formula. If it is not specified, the last part of each variable is taken as reference.

`B` It is an intenger and only works if `pres = TRUE` and `simplex_pvalues = "bootstrap"`. It indicates the number of bootstrap samples.

`seed` It is an integer and only works if `pres = TRUE` and `simplex_pvalues = "bootstrap"`. It indicates the seed for the sampling.

Details

Remark that (semi-)elasticities in models where the endogenous is compositional are different across statistical units.

Value

A matrix where each element (i,j) represents the computed p-value for the (semi-)elasticity between variables 'j' and 'i'.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

See Also

In the case where the endogenous variable is not compositional, it is possible to use the usual t-test, see `sd_se_x` and `impacts`.

Examples

```
# Load data
data(BDDSegX)

# Compute the matrix of p-values for the 3rd observation
p_values_se(BDDSegX, formula = cbind(S_A,S_B,S_C,S_D,S_E) ~
     PIB_Courant_t + HT_Gazole + PAC, observation=3, B=300)
```

reg *Compositional regression function.*

Description

This function performs the estimation of coefficients in models where compositional variables are involved and presents them in a nice way. This is on of the main functions of the package, second to `impacts()`.

Usage

```
reg(dataset, formula, transformation = "ILR", V = list(matrix()), R = list(NULL),
    pres = FALSE, simplex_pvalues = FALSE, B = 1000, seed = 31000)
```

Arguments

`dataset`	Dataset in data.frame format.
`formula`	Formula of the desired model. Compositional variables should be inserted as a concatenation of their parts thorugh 'cbind()' and standard ones, although multivariate, with a plus '+' sign between all of them.
`transformation`	This argument can be "ILR" or "ALR" and indicates which type of transformation is employed.
`V`	This argument is only used if `transformation="ILR"`. It is a list. Each element of V is a matrix that is desired to be employed in the ILR transformation. The variable for which the i-th matrix is used is the i-th compositional variable, in order of appearance in the formula. If no list V is determined, contrast matrices constructed using the Helmert transformation are used.
`R`	This argument is only used if `transformation="ALR"`. It is a list of integers. Each element of R designs which part of the compositional variable is desired to be taken in the ALR transformation as reference. The variable for which the i-th integer is used is the i-th compositional variable, in order of appearance in the formula. If it is not specified, the last part of each variable is taken as reference.
`pres`	It is a logical variable. `pres = TRUE` and the results of the regression are displayed in a nice way, instead of returning the named list of results that are used to compute impacts.
`simplex_pvalues`	It is a categorical variable. Only works if `pres = TRUE`. If `simplex_pvalues = "bootstrap"`, the usual significance stars next to the simplex coefficients are displayed. The number of stars are calculated via bootstrap through `p_values_coef()`.
`B`	It is an intenger and only works if `pres = TRUE` and `simplex_pvalues = "bootstrap"`. It indicates the number of bootstrap samples.
`seed`	It is an integer and only works if `pres = TRUE` and `simplex_pvalues = "bootstrap"`. It indicates the seed for the sampling.

Details

Results in the simplex do not depend on the chosen transformation, but they do in the coordinate space.

Value

If pres=FALSE nothing is returned, results are only printed.

If pres=TRUE, the output depends on the nature of the variables included in the model.

Thus, if there are some exogenous compositional variables and the endogenous one is also compositional, see `alr_xy_reg()` or `ilr_xy_reg()` functions, depending on the case. If there are some exogenous compositional variables, but the endogenous one is standard, see `alr_x_reg()` or `ilr_x_reg()`. Finally, if the unique compositional variable is the endogenus one, see `alr_y_reg()` or `ilr_y_reg()`.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Kim, K., & Timm, N. (2006). *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press.

Nguyen, T. H. A., Laurent, T., Thomas-Agnan, C., & Ruiz-Gazen, A. (2020). Analyzing the impacts of socio-economic factors on French departmental elections with CoDa methods. *Journal of Applied Statistics*, 1-17. doi: 10.1080/02664763.2020.1858274

Examples

```
# Load data
data(BDDSegX)

# Perfom the regression and just
# present the results (thus pres=TRUE),
# including p-values from bootstrap
reg(BDDSegX, formula =
      cbind(S_A,S_B,S_C,S_D,S_E) ~ PAC,
      pres=TRUE, simplex_pvalues = "bootstrap")
```

`sd_se_x` *S.d. for semi-elasticities with non-compositional Y.*

Description

This function computes the standard deviation for semi-elasticities where the endogenous variable is not compositional. The formula is derived directly from the reference: Var(s.e) = Var(Cb*) = CVar(b*)C', where C is any contrast matrix.

The results of this function are used in `impacts()` to calculate and display the p-values of the significance test.

Usage

```
sd_se_x(reg_results)
```

Arguments

`reg_results`	Named list that is the output returned by `reg()`.

Details

These semi-elasticities are not dependent on the statistical unit, contrary to what happens in the case with an endogenous compositional variable. Therefore, there is only one matrix of standard deviations.

This function is intended to be integrated in `impacts()`, but could be usable directly.

Value

A matrix where the element (i,j) indicates the standard deviation of the semi-elasticity between variables 'i' and 'j'.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Morais, J., Thomas-Agnan, C., & Simioni, M. (2018). Interpretation of explanatory variables impacts in compositional regression models. *Austrian Journal of Statistics*, **47**(5), 1-25. doi: 10.17713/ajs.v47i5.718

See Also

In the case where the endogenous variable is compositional, it is possible to compute p-values through bootstrapping, see `p_values_se`.

`share_ratio` *Share ratio.*

Description

This function computes the ratio of shares, which are used for interpretation, for models where the endogenous variable is the only compositional one.

Usage

```
share_ratio(results)
```

Arguments

`results`	Named list which is returneg by `reg()` with argument `pres=FALSE`.

Value

It is returned a list of matrices, one for each exogenous variable. The k-th matrix is composed by elements (i,j) which are computed as log(b_ki)/log(b_kj), being b_ij the (i,j) element of the matrix of coefficients.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

Morais, J., Thomas-Agnan, C., & Simioni, M. (2017). Interpreting the impact of explanatory variables in compositional models.

Examples

```
# Load data
data(BDDSegX)

# Obtain regression results
reg_results = reg(BDDSegX,
    formula = cbind(S_A,S_B,S_C,S_D,S_E) ~
    PIB_Courant_t + HT_Gazole + PAC)

# Compute the ratio of shares
share_ratio(reg_results)
```

`tidy` *Tidy.*

Description

Function to present results in a nicer way.

Usage

```
tidy(matrix, dig = 14)
```

Arguments

`matrix`	Matrix which contains the results.
`dig`	Number of total digits to be kept for each entry.

Details

This is an internal function that is very conditioned to the nature of results of this package.

Value

The same matrix is returned, but featuring a layout where the minus signs in one row are not below or above digits in another row and all entries have the same number of digits.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

`var2alr` *Transform every variable to ALR coordinates.*

Description

This function is an intermediate function that takes every variable to be used in an ALR regression and transform it to the coordinates, if the variable is not already in coordinates. Moreover, it bundles into a single matrix all the exogenous variables to be able to use it as an input for `mlm()` function.

Usage

```
var2alr(Y, X, Z = list(matrix()), R = list(matrix()), constant = TRUE)
```

Arguments

`Y`	Data from the endogenous variable. It is a DxN matrix where D is the number of parts of the compositional endogenous variable and N is the number of statistical units.
`X`	Data from exogenous variables which are compositional. It is a list of (P_k)xN matrices. The k-th matrix corresponds to the k-th compositional exogenous variable. P_k stands for the number of parts that features the k-th variable.
`Z`	Data from exogenous variables which are not compositional (therefore standard). It is a list of (Q_k)xN matrices. The k-th matrix corresponds to the k-th standard exogenous variable. Q_k stands for the number of components that features the k-th variable, which may be multivariate or not.
`R`	List of integers where the i-th element indicates which part of the i-th compostional variable is desired to be used as reference for the transformation. The order taken into account is the order of appearance of each compositional variable (from left to right) in the function. If `R` is not specified, the last part of each compositional variable is taken as reference.
`constant`	Logical variable. `TRUE` if an intercept is desired and `FALSE` otherwise.

Details

As mentioned, this is an internal function and it is not intended to be used directly.

Value

It returns a named list with the following elements:

`Y`	A matrix with the endogenous variable transformed into the ALR coordinates.
`X`	One matrix with the exogenous variables bundled (both compositional and non-compositional) in the ALR coordinates.
`R`	It is a list with integers that denote the references that are eventually used inside the function.
`constant`	A logical variable that returns TRUE if an intercept is considered and FALSE otherwise.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

`var2ilr` *Transform every variable to ILR coordinates.*

Description

This function is an intermediate function that takes every variable to be used in an ILR regression and transform it to the coordinates, if the variable is not already in coordinates. Moreover, it bundles into a single matrix all the exogenous variables to be able to use it as an input for `mlm()` function.

Usage

```
var2ilr(Y, X, Z = list(matrix()), V = list(matrix()), constant = TRUE)
```

Arguments

`Y`	Data from the endogenous variable. It is a DxN matrix where D is the number of parts of the compositional endogenous variable and N is the number of statistical units.
`X`	Data from exogenous variables which are compositional. It is a list of (P_k)xN matrices. The k-th matrix corresponds to the k-th compositional exogenous variable. P_k stands for the number of parts that features the k-th variable.
`Z`	Data from exogenous variables which are not compositional (therefore standard). It is a list of (Q_k)xN matrices. The k-th matrix corresponds to the k-th standard exogenous variable. Q_k stands for the number of components that features the k-th variable, which may be multivariate or not.
`V`	List where the i-th element indicates which contrast matrix is desired to be used as reference for the transformation. The order taken into account is the order of appearance of each compositional variable (from left to right) in the function. If `V` is not specified, `V_D(D)` is used for the endogenous variable and `V_D(Q_k)` is used for the k-th compositional exogenous variable.
`constant`	Logical variable. `TRUE` if an intercept is desired and `FALSE` otherwise.

Details

As mentioned, this is an internal function and it is not intended to be used directly.

Value

It returns a named list with the following elements:

`Y`	A matrix with the endogenous variable transformed into the ALR coordinates.
`X`	One matrix with the exogenous variables bundled (both compositional and non-compositional) in the ILR coordinates.
`V`	It is a list with the contrast matrices that are eventually used inside the function.
`constant`	A logical variable that returns TRUE if an intercept is considered and FALSE otherwise.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

V_D *Construction of the helmertian contrast matrices.*

Description

This function constructs the Dx(D-1) contrast matrix as shown in the equation (2.11) of the reference, following the helmertian approach. These matrices act as default contrast matrices in the ILR transformation in case no other are specified.

Usage

```
V_D(D)
```

Arguments

D	Number of rows of the desired contrast matrix, whcih is Dx(D-1).

Value

Returns the Dx(D-1) contrast matrix based on the helmertian approach.

Author(s)

Ivan Rendo Barreiro, Thibault Laurent and Christine Thomas-Agnan.

References

van den Boogaart, K. G., & Tolosana-Delgado, R. (2013). *Fundamental concepts of compositional data analysis. In Analyzing Compositional Data with R* (pp. 13-50). Springer, Berlin, Heidelberg.

Index