

Package: regressionworlddata (via r-universe)

August 20, 2024

Type Package

Title Regression analysis based on global datasets

Version 1.1.5

Date 2024-06-03

Maintainer Benjamin Leon Bodirsky <bodirsky@pik-potsdam.de>

Description Model estimates parameters of model functions.

License LGPL-3 | file LICENSE

Depends madrat(>= 1.28), magclass(>= 3.17), magpiesets(>= 0.27),
mrcommons, R(>= 2.10.0)

Imports boot, countrycode, graphics, lmtest, mrfaocore, nlstools,
RColorBrewer, sandwich, stats

Suggests covr, testthat

Encoding UTF-8

LazyData no

RoxygenNote 7.3.1

Repository <https://pik-piam.r-universe.dev>

RemoteUrl <https://github.com/pik-piam/regressionworlddata>

RemoteRef HEAD

RemoteSha 23ad0c5380400401fcb54a77633668f8ee9d84cb

Contents

regressionworlddata-package	2
calcCollectRegressionData	2
nlsAddLines	3
nlsregression	4
robust_vce	6
toolCollectRegressionVariables	7
toolRegression	7
toolRegressionTable	9

Index	11
--------------	-----------

regressionworldldata-package

Moinput Regression function library

Description

Package contains functions to estimate model parameters

Details

Package: regressionworldldata
Type: Package
Version: 0.1
Date: 2016-09-23
License: LGPL-3
LazyLoad: yes

Author(s)

Benjamin Leon Bodirsky, Antonia Walther

Maintainer: Benjamin Leon Bodirsky <bodirsky@pik-potsdam.de>

calcCollectRegressionData

calcCollectRegressionData

Description

collects regression data using un-converted raw data sources, and crops the data that only joint years and countries are selected.

Usage

```
calcCollectRegressionData(datasources)
```

Arguments

`datasources` All datasources that shall be returned. Due to the cropping of data which is not present in all datasources, reducing the number of datasources will increase the number of observations.

Value

List of magpie objects with results on country level, weight on country level, unit and description.

Author(s)

Benjamin Leon Bodirsky, Eleonora Martinelli, Abhijeet Mishra, Xiaoxi Wang

Examples

```
## Not run:
calcOutput("CollectRegressionData", aggregate=F)

## End(Not run)
```

nlsAddLines

nlsAddLines

Description

Adds lines of specific countries into the plot of the function nlsregression. nlsregression has to be based on magpie objects for x,y,weight

Usage

```
nlsAddLines(
  y,
  x,
  countries = 1:5,
  weight = NULL,
  x_log10 = FALSE,
  colors = "black",
  labels = TRUE
)
```

Arguments

y	magpie object with y values
x	magpie object with x values
countries	Choice of countries
weight	magpie object with weight
x_log10	same as in nlsregression
colors	colors of the lines
labels	If TRUE, the region, startyear and endyear will be plotted to each line.

Value

vector with ISO-countrycodes

Author(s)

Benjamin Leon Bodirsky

See Also

[nlsregression](#)

Examples

```
## Not run:
data(population_magpie)
nlsregression(y=population_magpie[,1],x=population_magpie[,2],
weight = population_magpie[,1],func = y~a*x+b)
nlsAddLines(y=population_magpie[,1],x=population_magpie[,2],
weight = population_magpie[,1],countries=1:3,colors=1:3)

## End(Not run)
```

nlsregression

nlsregression

Description

Creates regression parameter estimates and plots with any function you want that has no more than two independent variables

Usage

```
nlsregression(
  func,
  y,
  x,
  z = NULL,
  startvalues = NULL,
  weight = NULL,
  weighting = TRUE,
  xlab = NULL,
  ylab = "y",
  header = NULL,
  z_plot_lines = NULL,
  weightcolorpoints = TRUE,
  x_log10 = FALSE,
  toPlot = "all",
  plot_x_function = "ignore",
```

```

    regressioncolor = "blue",
    weight_threshold = NULL,
    crossvalid = NULL,
    ...
)

```

Arguments

func	function that shall be fitted. Function should contain the dependent variable y and the independent variable x, eventually a second independent variable z. All other unknowns are treated as parameters that are estimated.
y	dependent variable,vector
x	independent variable,vector
z	optional independent variable,vector
startvalues	the optimization algorithm may require starting values for the fitting procedure. provide them in a list with the parameter names: e.g. list(a=3,b=2)
weight	optional weight,vector
weighting	if weighting is TRUE, the fit will minimize the weighted residuals
xlab	name of x axis in plot
ylab	name of y axis in plot
header	plot function main argument
z_plot_lines	vector>1 of values for z you want to be plotted into the graph
weightcolorpoints	if TRUE, the points are clustered into three quantiles according to their weight and coloured lighter for low weights.
x_log10	allows log10 scale for X axis if set to TRUE. Only changes the picture, not the regression!
toPlot	"all", "frame" (axis etc), "observations" (points), "regressionline" (line), "infos" (parameters, R2)
plot_x_function	deprecated, please do not enter into function call.
regressioncolor	color of regression line and parameter text
weight_threshold	if numeric, all countries below this threshold will be excluded (e.g. to exclude minor islands)
crossvalid	vector with boolean values, indicating which data should be excluded from sampling and rather be used for validation
...	will be passed on to function nls

Value

A nice picture and regression parameters or eventually some errors.

Author(s)

Benjamin Leon Bodirsky, Susanne Rolinski, Xiaoxi Wang

Examples

```
## Not run:
x=1:10
y=(1:10)^2+1
z=c(10:1)

# one independent variable
nlsregression(func=y~a*x+b,y=y,x=x,startvalues=list(a=1,b=1))
# two independent variables
nlsregression(func=y~a*x^1.1+b*z+c*x,y=y,x=x,z=z,startvalues=list(a=1,b=1,c=0))
# no fit because residuals are zero (excluded from the nls makers due
  to statistical reasons)
nlsregression(func=y~x^a+b,y=y,x=x,z=z,startvalues=list(a=1,b=1,c=0))

DNase1 <- subset(DNase, Run == 1)
DNase1$sets<- c(rep(1,8),rep(2,8))
nlsregression(func=y~a*x+b,y=DNase1$density,x=DNase1$conc,startvalues=list(a=1,b=1))
nlsregression(func=y~a*x+b*z,y=DNase1$density,x=DNase1$conc,z=DNase1$sets,
startvalues=list(a=0.1344,b=0.2597))
nlsregression(func=y~a*x+b*z,y=DNase1$density,x=DNase1$conc,z=DNase1$sets,
startvalues=list(a=0.1344,b=0.2597),plot_x_function=log)

## End(Not run)
```

robust_vce

robust_vce

Description

returns robust var-cov estimate

Usage

```
robust_vce(x)
```

Arguments

x regression model

Value

a robust estimate of variance-covariance matrix and corresponding t-value and p-value for estimated coefficients

Author(s)

Xiaoxi Wang

`toolCollectRegressionVariables`
toolCollectRegressionVariables

Description

todo

Usage

```
toolCollectRegressionVariables(indicators)
```

Arguments

indicators todo

Value

todo

Author(s)

Benjamin Leon Bodirsky

`toolRegression` *toolRegression*

Description

Regression model for the correlation of a denominator and quotient to the GDP, allowing for an additional driver z next to income.

Usage

```
toolRegression(  
  denominator,  
  quotient = NULL,  
  func = y ~ (a * x)/(b + x),  
  x = "IHME_USD05_PPP_pc",  
  z = NULL,  
  ylab = NULL,  
  xlab = NULL,
```

```

data = NULL,
countries_nlsAddLines = NULL,
weight = "pop",
x_log10 = FALSE,
crossvalid_sample = NULL,
crossvalid_drawing = 1,
...
)

```

Arguments

denominator	denominator of the dependent variable that shall be estimated using the regression
quotient	quotient of the dependent variable that shall be estimated using the regression
func	functional relation for the regression, shall be in the format $y \sim f(x, \dots)$ with x being <code>gdp</code> , y being denominator/quotient, and $f()$ being any type of functional relationship. ... can include either z or parameters to be estimated.
x	independent variable, by default <code>income</code>
z	additional independent variable
<code>ylab</code>	name of y axis
<code>xlab</code>	name of x axis
<code>data</code>	data can be provided if Data shall not be derived by <code>mrcommons:::calcCollectFoodDemandRegressionData</code>
<code>countries_nlsAddLines</code>	the number of weightiest countries or the name of countries that shall be plotted by lines in the plot
<code>weight</code>	the weight
<code>x_log10</code>	passed on to <code>nlsregression()</code>
<code>crossvalid_sample</code>	sample name from <code>madrat</code> used for crossvalidation. Name is built as follows: <code>crossvalid_seedX_kY</code> X is the random seed, Y is the number of drawings. The combination of all drawings is the full sample.
<code>crossvalid_drawing</code>	selected drawing of k in <code>crossvalidsample</code>
...	further attributes that will be handed on to <code>nlsregression()</code> : An additional explanatory variable z can be added. A regression model has to be chosen. Startvalues can be predetermined.

Value

regression plot and the parameters from `nlsregression`

Author(s)

Antonia Walther, Benjamin Leon Bodirsky

See Also[calcOutput](#)**Examples**

```
## Not run:

toolRegression(denominator=livestock,
               func=y~(a*x)/(b+x),
               z=NULL,
               startvalues=list(a=1100,b=7770)
               )

toolRegression(denominator=findset("kap"),
               quotient=findset("kfo"),
               func=y~(a*x)/(b+x),
               z=NULL,
               startvalues=list(a=0.5,b=7770)
               )

## End(Not run)
```

toolRegressionTable *toolRegressionTable*

Description

creates Regression for selected options and saves calculated parametes inside the table.

Usage

```
toolRegressionTable(
  scenario = "SSP2",
  x = "IHME_USD05_PPP_pc",
  denominator = NA,
  z = NA,
  regression_database_file = "scenario_database_regressionworldldata.csv",
  quotient = "pop",
  start_1 = NA,
  start_2 = NA,
  start_3 = NA,
  start_4 = NA,
  start_5 = NA,
  start_6 = NA,
  return_value = FALSE
)
```

Arguments

scenario	vector. Default "SSP2". Can be "SSP1", "SSP2", "SSP3", "SSP4", "SSP5" or "mix" and describes the overall scenario of the projection.
x	Indep Var
denominator	vector. Default NA. Specific foddenominator share to make projection for.
z	other independent variables
regression_database_file	file with regressions to calculate
quotient	vector. Default is population ("pop")
start_1	Default NA. Startvalue for 1st parameter.
start_2	Default NA. Startvalue for 2nd parameter.
start_3	Default NA. Startvalue for 3rd parameter.
start_4	Default NA. Startvalue for 4th parameter.
start_5	Default NA. Startvalue for 5th parameter.
start_6	Default NA. Startvalue for 6th parameter.
return_value	Default to False. This is to stop printing the updated dataset on console. If you'd like to keep the updated dataset as an object, set this to true.

Value

data frame with additional rows containing parameters of newly calculated regression.

Author(s)

Abhijeet Mishra, Eleonora Martinelli

See Also

[toolRegression](#)

Index

`calcCollectRegressionData`, [2](#)
`calcOutput`, [9](#)

`nlsAddLines`, [3](#)
`nlsregression`, [4](#), [4](#)

`regressionworlddata`
 (`regressionworlddata-package`),
 [2](#)

`regressionworlddata-package`, [2](#)
`robust_vce`, [6](#)

`toolCollectRegressionVariables`, [7](#)
`toolRegression`, [7](#), [10](#)
`toolRegressionTable`, [9](#)