

Package ‘lagsarlmtree’

October 13, 2022

Title Spatial Lag Model Trees

Date 2019-04-06

Version 1.0-1

Description Model-based linear model trees adjusting for spatial correlation using a simultaneous autoregressive spatial lag, Wagner and Zeileis (2019) [doi:10.1111/geer.12146](https://doi.org/10.1111/geer.12146).

Depends R (>= 3.3.0), partykit, spatialreg

Imports Formula (>= 1.2-1)

License GPL-2 | GPL-3

NeedsCompilation no

Author Martin Wagner [aut],
Achim Zeileis [aut, cre] (<https://orcid.org/0000-0003-0918-3766>),
Roger Bivand [ctb] (<https://orcid.org/0000-0003-2392-6140>)

Maintainer Achim Zeileis <Achim.Zeileis@R-project.org>

Repository CRAN

Date/Publication 2019-04-08 08:36:49 UTC

R topics documented:

GrowthNUTS2	2
lagsarlmtree	4
WeightsNUTS2	6
Index	8

Description

Growth regression data for NUTS2 regions in the European Union.

Usage

```
data("GrowthNUTS2")
```

Format

A data frame containing 255 observations on 58 variables.

ggdpcap numeric. Average annual growth rate of real GDP per capita over the period 1995-2005.

accessair numeric. Measure for potential accessibility by air.

accessrail numeric. Measure for potential accessibility by rail.

accessroad numeric. Measure for potential accessibility by road.

airportdens numeric. Airport density (number of airports per sqkm).

airports factor. Number of airports.

arh0 numeric. Initial activity rate, highly educated.

arl0 numeric. Initial activity rate, low educated.

arm0 numeric. Initial activity rate, medium educated.

art0 numeric. Initial activity rate, total.

capital factor. Does the region host country capital city?

connectair numeric. Connectivity to comm. airports by car of the capital or centroid of region.

connectsea numeric. Connectivity to comm. seaports by car of the capital or centroid of region.

distcap numeric. Distance to capital city of respective country.

distde71 numeric. Distance to Frankfurt.

empdens0 numeric. Initial employment density.

ereh0 numeric. Initial employment rate, highly educated.

erel0 numeric. Initial employment rate, low educated.

erem0 numeric. Initial employment rate, medium educated.

eret0 numeric. Initial employment rate, total.

gdpcap0 numeric. Real GDP per capita in logs in 1995.

gpop numeric. Growth rate of population.

hazard numeric. Sum of all weighted hazard values.

hrstcore numeric. Human resources in science and technology (core).

intf numeric. Proportion of firms with own website regression.

outdens0 numeric. Initial output density.

popdens0 numeric. Initial population density.

raildens numeric. Rail density (length of railroad network in km per sqkm).

regboarder factor. Border region?

regcoast factor. Coastal region?

regobj1 factor. Is the region within an Objective 1 region?

regpent27 factor. Pentagon EU 27 region? (London, Paris, Munich, Milan, Hamburg.)

roaddens numeric. Road density (length of road network in km per sqkm).

seaports factor. Does the region have a seaport?

settl factor. Settlement structure.

shab0 numeric. Initial share of NACE A and B (Agriculture) in GVA.

shce0 numeric. Initial share of NACE C to E (Mining, Manufacturing and Energy) in GVA.

shgfcf numeric. Share of gross fixed capital formation in gross value added.

shjk0 numeric. Initial share of NACE J to K (Business services) in GVA.

shsh numeric. Share of highly educated in working age population.

shsl numeric. Share of low educated in working age population.

shlll numeric. Life long learning.

shsm numeric. Share of medium educated in working age population.

telf factor. A typology of estimated levels of business telecommunications access and uptake.

temp numeric. Extreme temperatures.

urh0 numeric. Initial unemployment rate, highly educated.

url0 numeric. Initial unemployment rate, low educated.

urm0 numeric. Initial unemployment rate, medium educated.

urt0 numeric. Initial unemployment rate, total.

country factor. Country within which the region is located.

cee factor. Is the region within a Central and Eastearn European country?

piigs factor. Is the region within a PIIGS country? (Portugal, Ireland, Italy, Greece, Spain.)

de factor. Is the region within Germany?

es factor. Is the region within Spain?

fr factor. Is the region within France?

it factor. Is the region within Italy?

pl factor. Is the region within Poland?

uk factor. Is the region within the United Kingdom?

References

Schneider U, Wagner M (2012). Catching Growth Determinants with the Adaptive Lasso. *German Economic Review*, **13**(1), 71-85. doi: [10.1111/j.14680475.2011.00541.x](https://doi.org/10.1111/j.14680475.2011.00541.x)

Examples

```
data("GrowthNUTS2")
summary(GrowthNUTS2)
```

lagsarlmtree

*Spatial Lag Model Trees***Description**

Model-based recursive partitioning based on linear regression adjusting for a (global) spatial simultaneous autoregressive lag.

Usage

```
lagsarlmtree(formula, data, listw = NULL, method = "eigen",
             zero.policy = NULL, interval = NULL, control = list(),
             rhowystart = NULL, abstol = 0.001, maxit = 100,
             dfsplit = TRUE, verbose = FALSE, plot = FALSE, ...)
```

Arguments

formula	formula specifying the response variable and regressors and partitioning variables, respectively. For details see below.
data	data.frame to be used for estimating the model tree.
listw	a weights object for the spatial lag part of the model.
method	"eigen" (default) - the Jacobian is computed as the product of $(1 - \rho \cdot \text{eigenvalue})$ using eigenw, and "spam" or "Matrix_J" for strictly symmetric weights lists of styles "B" and "C", or made symmetric by similarity (Ord, 1975, Appendix C) if possible for styles "W" and "S", using code from the spam or Matrix packages to calculate the determinant; "Matrix" and "spam_update" provide updating Cholesky decomposition methods; "LU" provides an alternative sparse matrix decomposition approach. In addition, there are "Chebyshev" and Monte Carlo "MC" approximate log-determinant methods; the Smirnov/Anselin (2009) trace approximation is available as "moments". Three methods: "SE_classic", "SE_whichMin", and "SE_interp" are provided experimentally, the first to attempt to emulate the behaviour of Spatial Econometrics toolbox ML fitting functions. All use grids of log determinant values, and the latter two attempt to ameliorate some features of "SE_classic".
zero.policy	default NULL, use global option value; if TRUE assign zero to the lagged value of zones without neighbours, if FALSE (default) assign NA - causing lagsarlm() to terminate with an error
interval	default is NULL, search interval for autoregressive parameter
control	list of extra control arguments - see lagsarlm
rhowystart	numeric. A vector of length nrow(data), to be used as an offset in estimation of the first tree. NULL by default, which results in an initialization with the root model (without partitioning).
abstol	numeric. The convergence criterion used for estimation of the model. When the difference in log-likelihoods of the model from two consecutive iterations is smaller than abstol, estimation of the model tree has converged.

<code>maxit</code>	numeric. The maximum number of iterations to be performed in estimation of the model tree.
<code>dfsplitt</code>	logical or numeric. <code>as.integer(dfsplitt)</code> is the degrees of freedom per selected split employed when extracting the log-likelihood.
<code>verbose</code>	Should the log-likelihood value of the estimated model be printed for every iteration of the estimation?
<code>plot</code>	Should the tree be plotted at every iteration of the estimation? Note that selecting this option slows down execution of the function.
<code>...</code>	Additional arguments to be passed to <code>lmtree()</code> . See mob_control documentation for details.

Details

Spatial lag trees learn a tree where each terminal node is associated with different regression coefficients while adjusting for a (global) spatial simultaneous autoregressive lag. This allows for detection of subgroup-specific coefficients with respect to selected covariates, while adjusting for spatial correlations in the data. The estimation algorithm iterates between (1) estimation of the tree given an offset of the spatial lag effect, and (2) estimation of the spatial lag model given the tree structure.

The code is still under development and might change in future versions.

Value

The function returns a list with the following objects:

<code>formula</code>	The formula as specified with the <code>formula</code> argument.
<code>call</code>	the matched call.
<code>tree</code>	The final <code>lmtree</code> .
<code>lagsarlm</code>	The final <code>lagsarlm</code> model.
<code>data</code>	The dataset specified with the <code>data</code> argument including added auxiliary variables <code>.rhowy</code> and <code>.tree</code> from the last iteration.
<code>nobs</code>	Number of observations.
<code>loglik</code>	The log-likelihood value of the last iteration.
<code>df</code>	Degrees of freedom.
<code>dfsplitt</code>	degrees of freedom per selected split as specified with the <code>dfsplitt</code> argument.
<code>iterations</code>	The number of iterations used to estimate the <code>lagsarlmtree</code> .
<code>maxit</code>	The maximum number of iterations specified with the <code>maxit</code> argument.
<code>rhowystart</code>	Offset in estimation of the first tree as specified in the <code>rhowystart</code> argument.
<code>abstol</code>	The prespecified value for the change in log-likelihood to evaluate convergence, as specified with the <code>abstol</code> argument.
<code>listw</code>	The <code>listw</code> object used.
<code>mob.control</code>	A list containing control parameters passed to <code>lmtree()</code> , as specified with <code>...</code>

References

Wagner M, Zeileis A (2019). Heterogeneity and Spatial Dependence of Regional Growth in the EU: A Recursive Partitioning Approach. *German Economic Review*, **20**(1), 67–82. doi: [10.1111/geer.12146](https://doi.org/10.1111/geer.12146) <https://eeecon.uibk.ac.at/~zeileis/papers/Wagner+Zeileis-2019.pdf>

See Also

[lm](#), [lagsarlm](#), [lmtree](#)

Examples

```
## data and spatial weights
data("GrowthNUTS2", package = "lagsarlm")
data("WeightsNUTS2", package = "lagsarlm")

## spatial lag model tree
system.time(tr <- lagsarlm(gdpcap ~ gdpcap0 + shgfcf + shsh + shsm |
  gdpcap0 + accessrail + accessroad + capital + regboarder + regcoast + regobj1 + cee + piigs,
  data = GrowthNUTS2, listw = WeightsNUTS2$invw,
  minsize = 12, alpha = 0.05))
print(tr)
plot(tr, tp_args = list(which = 1))

## query coefficients
coef(tr, model = "tree")
coef(tr, model = "rho")
coef(tr, model = "all")
system.time({
  ev <- eigenw(WeightsNUTS2$invw)
  tr1 <- lagsarlm(gdpcap ~ gdpcap0 + shgfcf + shsh + shsm |
    gdpcap0 + accessrail + accessroad + capital + regboarder + regcoast + regobj1 + cee + piigs,
    data = GrowthNUTS2, listw = WeightsNUTS2$invw, method = "eigen",
    control = list(pre_eig = ev), minsize = 12, alpha = 0.05)
})
coef(tr1, model = "rho")
```

WeightsNUTS2

Spatial Weights for European Union NUTS2 Regions

Description

Spatial weight matrices for NUTS2 regions in the European Union.

Usage

```
data("WeightsNUTS2")
```

Format

A list containing 40 listw weight matrices.

Source

Journal of Applied Econometrics Data Archive.

<http://qed.econ.queensu.ca/jae/2013-v28.4/cuaresma-feldkircher/>

References

Crespo Cuaresma J, Feldkircher M (2013). Spatial Filtering, Model Uncertainty and the Speed of Income Convergence in Europe. *Journal of Applied Econometrics*, **28**(4), 720-741. doi: [10.1002/jae.2277](https://doi.org/10.1002/jae.2277)

Index

* **datasets**

GrowthNUTS2, [2](#)

WeightsNUTS2, [6](#)

* **tree**

lagsarlmtree, [4](#)

coef.lagsarlmtree (lagsarlmtree), [4](#)

GrowthNUTS2, [2](#)

impacts.lagsarlmtree (lagsarlmtree), [4](#)

lagsarlm, [4](#), [6](#)

lagsarlmtree, [4](#)

lm, [6](#)

lmtree, [6](#)

logLik.lagsarlmtree (lagsarlmtree), [4](#)

mob_control, [5](#)

plot.lagsarlmtree (lagsarlmtree), [4](#)

predict.lagsarlmtree (lagsarlmtree), [4](#)

print.lagsarlmtree (lagsarlmtree), [4](#)

sctest.lagsarlmtree (lagsarlmtree), [4](#)

WeightsNUTS2, [6](#)