Package ‘hspm’

May 8, 2026

Type Package

Title Heterogeneous Spatial Models

Date 2023-03-07

Version 1.1

Maintainer Gianfranco Piras <gpiras@mac.com>

Description Spatial heterogeneity can be specified in various ways. 'hspm' is an ambitious project that aims at implementing various methodologies to control for heterogeneity in spatial models. The current version of 'hspm' deals with spatial and (non-spatial) regimes models. In particular, the package allows to estimate a general spatial regimes model with additional endogenous variables, specified in terms of a spatial lag of the dependent variable, the spatially lagged regressors, and, potentially, a spatially autocorrelated error term. Spatial regime models are estimated by instrumental variables and generalized methods of moments (see Arraiz et al., (2010) <doi:10.1111/j.1467-9787.2009.00618.x>, Bivand and Piras, (2015) <doi:10.18637/jss.v063.i18>, Drukker et al., (2013) <doi:10.1080/07474938.2013.741020>, Kelejian and Prucha, (2010) <doi:10.1016/j.jeconom.2009.10.025>).

Encoding UTF-8

LazyData true

RoxygenNote 7.2.2

Depends R (>= 4.0)

Imports Formula, sphet, stats, spdep, Matrix

Suggests splm

License GPL (>= 2)

URL https://github.com/gpiras/hspm

BugReports https://github.com/gpiras/hspm/issues

NeedsCompilation no

Author Gianfranco Piras [aut, cre] (ORCID: <https://orcid.org/0000-0003-0225-6061>), Mauricio Sarrias [aut] (ORCID: <https://orcid.org/0000-0001-5932-4817>)

Repository CRAN

Date/Publication 2023-03-08 08:40:02 UTC

Contents

`baltim` *Baltimore house sales prices and hedonics*

Description

A dataset containing the prices and other attributes of 211 dwelling in Baltimore, MD

Usage

```
baltim
```

Format

A data frame with 211 rows and 17 variables:

STATION ID variable
PRICE sales price, in 1,000 US dollars (MLS)
NROOM number of rooms
DWELL 1 if detached unit, 0 otherwise
NBATH number of bathrooms
PATIO 1 if patio, 0 otherwise
FIREPL 1 if fireplace, 0 otherwise
AC 1 if air conditioning, 0 otherwise
BMENT 1 if basement, 0 otherwise
NSTOR number of stores
GAR number of car space in garage, (0 = no garage)
AGE age of dwellings in years
CITCOU 1 if dwelling is in Baltimore County, 0 otherwise
LOTSZ lot size in hundreds of square feet
SQFT interior living space in hundreds of square feet
X X coordinate on the Maryland grid
Y Y coordinate on the Maryland grid

Source

https://geodacenter.github.io/data-and-lab/

ivregimes *Estimation of regime models with endogenous variables*

Description

The function `ivregimes` deals with the estimation of regime models. Most of the times the variable identifying the regimes reveals some spatial aspects of the data (e.g., administrative boundaries). The model includes exogenous as well as endogenous variables among the regressors.

Usage

```
ivregimes(formula, data, rgv = NULL, vc = c("homoskedastic", "robust", "OGMM"))
```

Arguments

`formula`	a symbolic description of the model of the form `y ~ x_f | x_v | h_f | h_v` where `y` is the dependent variable, `x_f` are the regressors that do not vary by regimes, `x_v` are the regressors that vary by regimes, `h_f` are the fixed instruments and `h_v` are the instruments that vary by regimes.
`data`	the data of class `data.frame`.
`rgv`	an object of class `formula` to identify the regime variables
`vc`	one of `c("homoskedastic", "robust", "OGMM")`. If `"OGMM"` an optimal weighted GMM is used to estimate the VC matrix.

Details

The basic (non spatial) model with endogenous variables can be written in a general way as:

$$y = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + X\beta + \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + Y\pi + \varepsilon$$

where $y = [y_1', y_2']'$, and the $n_1 \times 1$ vector y_1 contains the observations on the dependent variable for the first regime, and the $n_2 \times 1$ vector y_2 (with $n_1 + n_2 = n$) contains the observations on the dependent variable for the second regime. The $n_1 \times k$ matrix X_1 and the $n_2 \times k$ matrix X_2 are blocks of a block diagonal matrix, the vectors of parameters β_1 and β_2 have dimension $k_1 \times 1$ and $k_2 \times 1$, respectively, X is the $n \times p$ matrix of regressors that do not vary by regime, β is a $p \times 1$ vector of parameters. The three matrices Y_1 $(n_1 \times q)$, Y_2 $(n_2 \times q)$ and Y $(n \times r)$ with corresponding vectors of parameters π_1, π_2 and π, contain the endogenous variables. Finally, $\varepsilon = [\varepsilon_1', \varepsilon_2']'$ is the $n \times 1$ vector of innovations. The model is estimated by two stage least square. In particular:

- If `vc = "homoskedastic"`, the variance-covariance matrix is estimated by $\sigma^2(\hat{Z}'\hat{Z})^{-1}$, where $\hat{Z} = PZ$, $P = H(H'H)^{-1}H'$, H is the matrix of instruments, and Z is the matrix of all exogenous and endogenous variables in the model.
- If `vc = "robust"`, the variance-covariance matrix is estimated by $(\hat{Z}'\hat{Z})^{-1}(\hat{Z}'\hat{\Sigma}\hat{Z})(\hat{Z}'\hat{Z})^{-1}$, where $\hat{\Sigma}$ is a diagonal matrix with diagonal elements $\hat{\sigma}_i$, for $i = 1, ..., n$.

- Finally, if vc = "OGMM", the model is estimated in two steps. In the first step, the model is estimated by 2SLS yielding the residuals $\hat{\varepsilon}$. With the residuals, the diagonal matrix $\hat{\Sigma}$ is estimated and is used to construct the matrix $\hat{S} = H'\hat{\Sigma}H$. Then $\eta_{OWGMM} = (Z'H\hat{S}^{-1}H'Z)^{-1}Z'H\hat{S}^{-1}H'y$, where η_{OWGMM} is the vector of all the parameters in the model, The variance-covariance matrix is: $n(Z'H\hat{S}^{-1}H'Z)^{-1}$.

Value

An object of class ivregimes. A list of five elements. The first element of the list contains the estimation results. The other elements are needed for printing the results.

Author(s)

Gianfranco Piras and Mauricio Sarrias

Examples

```
data("natreg")
form  <- HR90  ~ 0 | MA90 + PS90 + RD90 + UE90 | 0 | MA90 + PS90 + RD90 + FH90 + FP89 + GI89
split  <- ~ REGIONS
mod <- ivregimes(formula = form, data = natreg, rgv = split, vc = "robust")
summary(mod)
mod1 <- ivregimes(formula = form, data = natreg, rgv = split, vc = "OGMM")
summary(mod1)
form1  <- HR90  ~ MA90 + PS90 |  RD90 + UE90 -1 | MA90 + PS90 | RD90 + FH90 + FP89 + GI89 -1
mod2 <- ivregimes(formula = form1, data = natreg, rgv = split, vc = "homoskedastic")
summary(mod2)
```

---

natreg *US Counties Homicides data*

Description

Continental U.S. counties data for homicides and selected socio-economic characteristics. Data for four decennial census years: 1960, 1970, 1980 and 1990.

Usage

```
natreg
```

Format

A data frame with 3085 rows and 73 variables

REGIONS Regions of the US

NOSOUTH Counties not in the south

POLY_ID Poligon id

NAME Counties names

STATE_NAME State name

STATE_FIPS FIPS code for the state

CNTY_FIPS FIPS code for the county

FIPS state and county FIPS code

STFIPS FIPS code for the state

COFIPS FIPS code for the county

FIPSNO state + county FIPS code

SOUTH dummy indicator: 1 if the county is in the southern US

HR60 homicide rate per 100,000 in 1960

HR70 homicide rate per 100,000 in 1970

HR80 homicide rate per 100,000 in 1980

HR90 homicide rate per 100,000 in 1990

HC60 homicide count, three year average centered on 1960

HC70 homicide count, three year average centered on 1970

HC80 homicide count, three year average centered on 1980

HC90 homicide count, three year average centered on 1990

PO60 county population in 1960

PO70 county population in 1970

PO80 county population in 1980

PO90 county population in 1990

RD60 resource deprivation in 1960

RD70 resource deprivation in 1970

RD80 resource deprivation in 1980

RD90 resource deprivation in 1990

PS60 population structure in 1960

PS70 population structure in 1970

PS80 population structure in 1980

PS90 population structure in 1990

UE60 unemployment rate in 1960

UE70 unemployment rate in 1970

UE80 unemployment rate in 1980

UE90 unemployment rate in 1990

DV60 divorce rate in 1960: pct. males over 14 divorced

DV70 divorce rate in 1970: pct. males over 14 divorced

DV80 divorce rate in 1980: pct. males over 14 divorced

DV90 divorce rate in 1990: pct. males over 14 divorced

MA60 median age in 1960

MA70 median age in 1970

MA80 median age in 1980

MA90 median age in 1990

POL60 log of population in 1960

POL70 log of population in 1970

POL80 log of population in 1980

POL90 log of population in 1990

DNL60 log of population density in 1960

DNL70 log of population density in 1970

DNL80 log of population density in 1980

DNL90 log of population density in 1990

MFIL59 log of median family income in 1959

MFIL69 log of median family income in 1969

MFIL79 log of median family income in 1979

MFIL89 log of median family income in 1989

FP59 pct. families below poverty in 1959

FP69 pct. families below poverty in 1969

FP79 pct. families below poverty in 1979

FP89 pct. families below poverty in 1989

BLK60 pct. black in 1960

BLK70 pct. black in 1970

BLK80 pct. black in 1980

BLK90 pct. black in 1990

GI59 Gini index of family income inequality in 1959

GI69 Gini index of family income inequality in 1969

GI79 Gini index of family income inequality in 1979

GI89 Gini index of family income inequality in 1989

FH60 pct. female headed households in 1960

FH70 pct. female headed households in 1970

FH80 pct. female headed households in 1980

FH90 pct. female headed households in 1990

West West regional dummy

Source

https://geodacenter.github.io/data-and-lab/

regimes *Estimation of regimes models*

Description

The function `regimes` deals with the estimation of regime models. Most of the times the variable identifying the regimes reveals some spatial aspects of the data (e.g., administrative boundaries).

Usage

```
regimes(formula, data, rgv = NULL, vc = c("homoskedastic", "groupwise"))
```

Arguments

`formula`	a symbolic description of the model of the form `y ~ x_f | x_v` where `y` is the dependent variable, `x_f` are the regressors that do not vary by regimes and `x_v` are the regressors that vary by regimes
`data`	the data of class `data.frame`.
`rgv`	an object of class `formula` to identify the regime variables
`vc`	one of `c("homoskedastic", "groupwise")`. If `groupwise`, the model VC matrix is estimated by weighted least square.

Details

For convenience and without loss of generality, we assume the presence of only two regimes. In this case, the basic (non-spatial) is:

$$y = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + X\beta + \varepsilon$$

where $y = [y_1', y_2']'$, and the $n_1 \times 1$ vector y_1 contains the observations on the dependent variable for the first regime, and the $n_2 \times 1$ vector y_2 (with $n_1 + n_2 = n$) contains the observations on the dependent variable for the second regime. The $n_1 \times k$ matrix X_1 and the $n_2 \times k$ matrix X_2 are blocks of a block diagonal matrix, the vectors of parameters β_1 and β_2 have dimension $k_1 \times 1$ and $k_2 \times 1$, respectively, X is the $n \times p$ matrix of regressors that do not vary by regime, β is a $p \times 1$ vector of parameters and $\varepsilon = [\varepsilon_1', \varepsilon_2']'$ is the $n \times 1$ vector of innovations.

- If `vc = "homoskedastic"`, the model is estimated by OLS.
- If `vc = "groupwise"`, the model is estimated in two steps. In the first step, the model is estimated by OLS. In the second step, the inverse of the (groupwise) residuals from the first step are employed as weights in a weighted least square procedure.

Value

An object of class `lm` and `spregimes`.

Author(s)

Gianfranco Piras and Mauricio Sarrias

Examples

```
data("baltim")
form   <- PRICE  ~ NROOM + NBATH + PATIO + FIREPL + AC + GAR + AGE + LOTSZ + SQFT
split  <- ~ CITCOU
mod <- regimes(formula = form, data = baltim, rgv = split, vc = "groupwise")
summary(mod)
form <- PRICE  ~ AC + AGE + NROOM + PATIO + FIREPL + SQFT | NBATH + GAR + LOTSZ - 1
mod <- regimes(form, baltim, split, vc = "homoskedastic")
summary(mod)
```

`spregimes` *Estimation of spatial regimes models*

Description

The function `spregimes` deals with the estimation of spatial regimes models. This is a general function that allows the estimation of various spatial specifications, including the spatial lag regimes model, the spatial error regimes model, and the spatial SARAR regimes model. Since the estimation is based on generalized method of moments (GMM), endogenous variables can be included. For further information on estimation, see details.

Usage

```
spregimes(
  formula,
  data = list(),
  model = c("sarar", "lag", "error", "ols"),
  listw,
  wy_rg = FALSE,
  weps_rg = FALSE,
  initial.value = NULL,
  rgv = NULL,
  het = FALSE,
  verbose = FALSE,
  control = list()
)

## S3 method for class 'spregimes'
coef(object, ...)

## S3 method for class 'spregimes'
```

```
vcov(object, ...)

## S3 method for class 'spregimes'
print(x, digits = max(3, getOption("digits") - 3), ...)

## S3 method for class 'spregimes'
summary(object, ...)

## S3 method for class 'summary.spregimes'
print(x, digits = max(5, getOption("digits") - 3), ...)

## S3 method for class 'spregimes'
residuals(object, ...)

## S3 method for class 'spregimes'
fitted(object, ...)
```

Arguments

formula	a symbolic description of the model of the form `y ~ x_f | x_v | wx | h_f | h_v | wh` where `y` is the dependent variable, `x_f` are the regressors that do not vary by regimes, `x_v` are the regressors that vary by regimes, `wx` are the spatially lagged regressors, `h_f` are the instruments that do not vary by regimes, `h_v` are the instruments that vary by regimes, `wh` are the spatially lagged instruments.
data	the data of class `data.frame`.
model	should be one of `c("sarar", "lag", "error", "ols")`
listw	a spatial weighting matrix of class `listw`, `matrix` or `Matrix`
wy_rg	default `wy_rg = FALSE`, the lagged dependent variable does not vary by regime (see details)
weps_rg	default `weps_rg = FALSE`, if `TRUE` the spatial error term varies by regimes (see details)
initial.value	initial value for the spatial error parameter
rgv	an object of class `formula` to identify the regime variables
het	heteroskedastic variance-covariance matrix
verbose	print a trace of the optimization
control	select arguments for the optimization
object	an object of class spregimes
...	additional arguments
x	an object of class spregimes
digits	number of digits

Details

The function `spregimes` is a wrapper that allows the estimation of a general spatial regimes model. For convenience and without loss of generality, we assume the presence of only two regimes. In this case the general model can be written as:

$$
\begin{aligned}
y = & W \begin{bmatrix} y_1 & 0 \\ 0 & y_2 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} + \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + X\beta + \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + Y\pi + \\
& W \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + WX\delta + W \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + WY\theta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}
\end{aligned}
$$

where

$$
\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = W \begin{bmatrix} \varepsilon_1 & 0 \\ 0 & \varepsilon_2 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} + u
$$

The model includes the spatial lag of the dependent variable, the spatial lag of the regressors, the spatial lag of the errors and, possibly, additional endogenous variables. The function `spregimes` estimates all of the nested specifications deriving from this model. There are, however, some restrictions. For example, if `weps_rg` is set to TRUE, all the regressors in the model should also vary by regimes. The estimation of the different models relies heavily on code available from the package **sphet**.

1. For the spatial lag (or Durbin) regimes model (i.e, when ρ_1 and ρ_2 are zero), an instrumental variable procedure is adopted, where the matrix of instruments is formed by the spatial lags of the exogenous variables and the additional instruments included in the `formula`. A robust estimation of the variance-covariance matrix can be obtained by setting `het = TRUE`.
2. For the spatial error regime models (i.e, when λ_1 and λ_2 are zero), the spatial coefficient(s) are estimated with the GMM procedure described in Kelejian and Prucha (2010) and Drukker et al., (2013). The difference between Kelejian and Prucha (2010) and Drukker et al., (2013), is that the former assume heteroskedastic innovations (`het = TRUE`), while the latter does not (`het = FALSE`).
3. For the SARAR regimes model, the estimation procedure alternates a series of IV and GMM steps. The variance-covariance can be estimated assuming that the innovations are homoskedastic (`het = FALSE`) as well as heteroskedastic (`het = TRUE`).

Value

An object of class “`spregimes`”

Author(s)

Gianfranco Piras and Mauricio Sarrias

References

Arraiz, I. and Drukker, M.D. and Kelejian, H.H. and Prucha, I.R. (2010) A spatial Cliff-Ord-type Model with Heteroskedastic Innovations: Small and Large Sample Results, *Journal of Regional Sciences*, **50**, pages 592–614.

Drukker, D.M. and Egger, P. and Prucha, I.R. (2013) On Two-step Estimation of a Spatial Auto regressive Model with Autoregressive Disturbances and Endogenous Regressors, *Econometric Review*, **32**, pages 686–733.

Kelejian, H.H. and Prucha, I.R. (2010) Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances, *Journal of Econometrics*, **157**, pages 53–67.

Gianfranco Piras (2010). sphet: Spatial Models with Heteroskedastic Innovations in R. *Journal of Statistical Software*, 35(1), 1-21. doi:10.18637/jss.v035.i01.

Roger Bivand, Gianfranco Piras (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, 63(18), 1-36. doi:10.18637/jss.v063.i18.

Gianfranco Piras, Paolo Postiglione (2022). A deeper look at impacts in spatial Durbin model with sphet. *Geographical Analysis*, 54(3), 664-684. https://onlinelibrary.wiley.com/doi/10.1111/gean.12318

Luc Anselin, Sergio J. Rey (2014). *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySal.* GeoDa Press LLC.

Examples

```
data("natreg")
data("ws_6")

form <-  HR90  ~ 0 | MA90 + PS90 +
RD90 + UE90 | 0 | 0 | MA90 + PS90 +
RD90 + FH90 + FP89 + GI89 | 0

form1 <-  HR90  ~ MA90 -1 |  PS90 +
RD90 + UE90 | 0 | MA90 -1 |  PS90 +
RD90 + FH90 + FP89 + GI89 | 0

form2 <-  HR90  ~ MA90 -1 |  PS90 +
RD90 + UE90 | MA90 | MA90 -1 |  PS90 +
RD90 + FH90 + FP89 + GI89 | 0

form3 <-  HR90  ~ MA90 -1 |  PS90 +
RD90 + UE90 | MA90 | MA90 -1 |  PS90 +
RD90 + FH90 + FP89 + GI89 | GI89

form4 <-  HR90  ~ MA90 -1 |  PS90 +
RD90 + UE90 | MA90 + RD90 | MA90 -1 |  PS90 +
RD90 + FH90 + FP89 + GI89 | GI89


split  <- ~ REGIONS

###################################################
# Linear model with regimes and lagged regressors #
###################################################
mod <- spregimes(formula = form2, data = natreg,
rgv = split, listw = ws_6, model = "ols")
summary(mod)

mod1 <- spregimes(formula = form3, data = natreg,
rgv = split, listw = ws_6, model = "ols")
```

```
summary(mod1)

mod2 <- spregimes(formula = form4, data = natreg,
rgv = split, listw = ws_6, model = "ols")
summary(mod2)


################################
# Spatial Error regimes model #
################################
mod <- spregimes(formula = form, data = natreg,
rgv = split, listw = ws_6, model = "error", het = TRUE)
summary(mod)
mod1 <- spregimes(formula = form, data = natreg,
rgv = split, listw = ws_6, model = "error",
weps_rg = TRUE, het = TRUE)
summary(mod1)
mod2 <- spregimes(formula = form1, data = natreg,
rgv = split, listw = ws_6, model = "error", het = TRUE)
summary(mod2)

################################
#  Spatial Lag regimes model  #
################################
mod4 <- spregimes(formula = form, data = natreg,
rgv = split, listw = ws_6, model = "lag",
het = TRUE, wy_rg = TRUE)
summary(mod4)
mod5 <- spregimes(formula = form1, data = natreg,
rgv = split, listw = ws_6, model = "lag",
het = TRUE, wy_rg = TRUE)
summary(mod5)

################################
# Spatial SARAR regimes model #
################################
mod6 <- spregimes(formula = form, data = natreg,
rgv = split, listw = ws_6, model = "sarar",
het = TRUE, wy_rg = TRUE, weps_rg = TRUE)
summary(mod6)
mod7 <- spregimes(formula = form, data = natreg,
rgv = split, listw = ws_6, model = "sarar",
het = TRUE, wy_rg = FALSE, weps_rg = FALSE)
summary(mod7)
mod8 <- spregimes(formula = form1, data = natreg,
rgv = split, listw = ws_6, model = "sarar",
het = TRUE, wy_rg = TRUE, weps_rg = FALSE)
summary(mod8)
```

---

ws_6 *Spatial weighting matrix for the US Counties Homicides data*

Description

ws_6 is a spatial weights matrix based on the 6 nearest neighbors for the Continental U.S. counties data for homicides

Usage

```
ws_6
```

Format

A spatial weighting matrix of class `Matrix`

Source

https://geodacenter.github.io/data-and-lab/

Index