Journal of Spatial Econometrics (2023) 4:4
https://doi.org/10.1007/s43071-023-00034-1

ORIGINAL PAPER

Heterogeneous spatial models in R: spatial regimes models

Gianfranco Piras[1,2] · Mauricio Sarrias[3]

Received: 15 March 2023 / Accepted: 26 June 2023

Abstract
This paper presents the progress made so far in the development of the R package hspm. The package hspm aims at implementing a variety of models and methods to control for heterogeneity in spatial models. Spatial heterogeneity can be specified in different ways, ranging from exogenous (or endogenous) spatial regimes models, to models with coefficients that potentially vary for each observations (i.e., continuous heterogeneity). We focus on a few R functions that allow for the estimation of a general spatial regimes model, as well as all of the nested specifications deriving from it. The models are estimated by instrumental variables and generalized method of moments techniques.

Keywords Spatial model · Heterogeneity · GMM · R

JEL Classification C21 · C87 · C88

1 Introduction

Spatial effects are generally divided into two different categories: spatial dependence and spatial heterogeneity (Anselin 1988). While cross-sectional dependence has to do with correlation between spatial units, spatial heterogeneity consists of instabilities over space that are generally reflected by variations across individual units (Anselin 2010).

In practice there are various ways of tackling unobserved heterogeneity, such as controlling for spatial heteroscedasticity (Kelejian and Prucha 2007), spatial regimes

✉ Gianfranco Piras
gpiras@mac.com

Mauricio Sarrias
mauricio.sarrias@utalca.cl

1 The Catholic University of America, Washington, USA

2 University of Chieti – Pescara, Chieti, Italy

3 Facultad de Economía y Negocios, Universidad de Talca, Talca, Chile

models (Anselin 1988; Anselin and Rey 2014), geographically weighted regressions (Fotheringham et al. 1998, 2002), and multilevel (or hierarchical) models (Arcaya et al. 2012), among others.[1]

An interesting distinction between discrete and continuous spatial heterogeneity has been made by Anselin and Amaral (2021). From a discrete perspective, they argue that spatial regimes models are the most common way of dealing with spatial heterogeneity. In a nutshell, spatial regimes models are a class of models whose coefficients may vary across space. The term regimes indicates that the observations are grouped according to some criteria that relates to space. Interestingly, Anselin and Amaral (2021) point out that, even if the estimation of spatial regimes regressions is well established, the identification of the regimes still remains a subject for investigation. Additionally, they acknowledge the existence of three approaches to identify the regimes. The first approach is based on exogenous regimes (e.g., determined through administrative boundaries); the second is when the regimes result from a data-driven procedure (e.g., observation are aggregated using some clustering method); and the last one corresponds to a situation where the coefficients and the regimes are jointly determined.

From an empirical perspective, attempts to consider spatial heterogeneity in model specification have mostly, but not exclusively, focused on economic geography and regional sciences. This is verified by the special attention that local labor markets (Huiban et al. 2004; Longhi and Nijkamp 2005; Melo et al. 2012), and regional economic convergence (Rey and Janikas 2005; Ramajo et al. 2008; Ertur et al. 2006) have received over the years. However, spatial heterogeneity has gained an increasing interest also in other disciplines, such as quantitative geography (Song et al. 2020; Georganos et al. 2021; Shu et al. 2019), urban growth (Zhai et al. 2021), urban sprawl (Deng et al. 2020; Irwin and Bockstael 2007), geology (de Marsily et al. 2005), ecology and evolution (Vinatier et al. 2011), epidemiology (Thomas et al. 2020), physics and air pollution (He et al. 2022), among others.

From a software availability perspective, spatial models to control for spatial dependence are well established.[2] Code dealing with spatial heterogeneity is relatively sparse but also long-established.[3] In this scenario, **hspm** is an ambitious project

[1] More recently, some authors focused on estimation and inference for spatial models with (continuous) heterogeneous coefficients in the context of spatial panel models (see, for example, Aquaro et al. 2020; Chen et al. 2022; LeSage and Chih 2018).

[2] There are various packages in R (R Development Core Team 2012), such as **spatialreg** (Bivand and Piras 2015; Bivand et al. 2021), **sphet** (Piras 2010; Piras and Postiglione 2022), **spldv** (Sarrias and Piras 2022) and **splm** (Millo and Piras 2012), among others, as well as in other software environment, such as the Spatial Econometrics toolbox in MATLAB (MATLAB 2011), the Python (van Rossum 1995) spatial analysis library **PySAL** (Anselin and Rey 2014), and Stata (StataCorp 2007).

[3] The packages in R mostly deal with geographically spatial regression (GWR), such as, **gwrr** (Wheeler 2022) **spgwr** (Bivand and Yu 2022), **mgwrsar** (Geniaux and Martinetti 2017), **GWmodel** (Gollini et al. 2015). Furthermore, **varycoef** (Dambon et al. 2021a, b) and **spBayes** (Finley et al. 2007, 2015) provide implementation of spatially varying coefficients models which may be preferred to GWR models as having proper statistical foundation. In the **PySAL** library developed in Python there is code dealing with spatial regime models. In Sect. 1 we compare our implementation with the one in **PySAL**.

that aims at developing and implementing various methodology to control for heterogeneity in spatial models. This article presents the methodological innovations that have been made so far dealing with spatial and (non spatial) regimes models. In particular, we present R functions that allow for the estimation of a general spatial regimes model, as well as all of the nested specifications deriving from it. The models are estimated by instrumental variables (IV) and generalized method of moments (GMM) techniques.

The rest of this paper is a mere description of the package functionality to get the readers to familiarize with the different functions contained in it. In particular, Sect. 2 introduces the two data sets that we use throughout the paper: the first is based on a housing price model in the city of Baltimore; the second contains county level data for homicides and selected socio-economic characteristics for the continental United States. The difference between these two data sets is that the second one suffers from endogeneity and requires instrumental variables methods implemented in **hspm**. In Sect. 3 we introduce the function `regimes` which is the basic function to deal with (non-spatial) regimes models. Section 4 is devoted to the illustration of the function `ivregimes` that allows for endogenous variables in a non-spatial context. The function `spregimes` is presented in Sect. 5. `spregimes` is a wrapper function that allows to estimate a regimes model with a spatial lag of the dependent variable, the spatial lag of (part of) the regressors, a spatially lagged error term and additional (other than the spatial lag) endogenous variables. As we will see, `spregimes` also allows to estimate all of the nested specifications included in this general model. In Sect. 6, we explain why **hspm** does not calculate the impacts measures put forth by LeSage and Pace (2009), and we show a simple way to deal with those impacts in a special case (i.e., when the spatial weighting matrix is block diagonal). Section 7 draws some conclusions and gives indications for future developments of the package. Finally, "Appendix A" compares our implementation with code available from **PySAL** library (Rey et al. 2022; Rey and Anselin 2007, 2010) developed in Python (van Rossum 1995).

2 Data sets

To illustrate the capabilities of **hspm** we make use of two data sets: `baltimore` and `natreg`.[4]

[4] We include the data and the spatial weighting matrices in the package. The original data are available at https://geodacenter.github.io/data-and-lab//. Anselin and Rey (2014) use the same data in Chapters 12 and 13.

Baltimore

The Baltimore data set on housing price (Dubir 1992) contains many standard factors to explain the price of a dwelling (PRICE): the number of rooms (NROOM), the number of bathrooms (NBATH), the age of the construction (AGE), the size of the lot (LOTSZ), the number of car space in a garage (GAR), and the square footage of the house (SQFT). Additional dummy variables are included to check whether the house has a patio (PATIO), a fireplace (FIREPL), and air conditioning (AC). The variable employed to identify the regimes is a binary equal to one if the dwelling is situated in Baltimore County and zero otherwise (CITCOU).

The following code loads the data and creates the spatial weighting matrix (of class `listw`) using a binary contiguity criterion.

```
R> library("hspm")
R> library("spdep")
R> data("baltim")
R> nbB <- read.gal(system.file("extdata",
+                              "baltimore.gal",
+                              package = "hspm"))
R> wlis <- nb2listw(nbB, style = "W")
```

natreg

The data in `natreg` contains information on homicides and selected socio-economic characteristics for the (continental) counties in the U.S., for four decennial census years, last of which is 1990 (Messner et al. 2000; Baller et al. 2001). Specifically, the dependent variable is the homicide rate in 1990 (HR90). Among the regressors we include median age (MA90), population structure (PS90), resource deprivation (RD90), and the, potentially endogenous, unemployment rate (UE90).[5] The instruments consist of three variables: percentage of female headed households (FH90), percentage of families below poverty (FP89), and the Gini index of family income inequality (GI89).[6] The regimes identifier is the variable REGIONS that divides the counties in three regions: south, west, and other (not south or west).

The following code loads the data and the spatial weighting matrix of class `Matrix` (Bates et al. 2022) based on the six nearest neighbors criteria:

```
R> data("natreg")
R> data("ws_6")
```

[5] The variables RD90 and PS90 were created from other indicators in a principal component analysis.

[6] Percentage of families below poverty and Gini index are based on 1989 figures.

3 The basic (non spatial) model and the function regimes

3.1 The basic (non spatial) model

For convenience and without loss of generality, we assume the presence of only two regimes (i.e., $j = 1, 2$). The basic (non spatial) model can be written in a general way as:

$$y = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + X\beta + \varepsilon, \tag{1}$$

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 of regressors, the vectors of parameters β_1 and β_2 have dimensions $k_1 \times 1$ 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-dimensional vector of regression disturbances.[7] Even though this is not a "traditional" spatial model, spatial heterogeneity is taken into account by considering a regimes variable that is revealing some spatial aspects of the data. The model in Eq. (1) can be estimated by OLS after reorganizing the data according to Eq. (1).[8]

3.2 The function regimes

The function `regimes` has four arguments: `formula`, `data`, `rgv`, and `vc`. The right hand side of the `formula` can be of different lengths. If the length is one, it is assumed that all coefficients are different by regimes. When the length of the formula is two, the variables in the first part are kept constant, while those in the second part are different by regimes.

The argument `rgv` is a `formula` that indicates the regimes variable. The are two options to estimate the variance-covariance matrix of the estimated coefficients: `"groupwise"` and `"homoskedastic"`: If `vc` is set to `"groupwise"`, the model is estimated according to a feasible generalized least squares procedure.[9]

In the example below, all the regressors vary by regimes and the `vc` argument is set to `"homoskedastic"`:

[7] Anselin and Rey (2014) refer to Eq. (1) as the hybrid model.

[8] The reorganization of the data is dealt internally by the function `regimes` that is introduced in Sect. 3.2.

[9] The feasible generalized least squares corresponds to a weighted least squares estimator where the weights for each regime are calculated as the sum of the squared residuals divided by the corresponding degrees of freedom (see Anselin and Rey 2014, for further details.

```
R> form_ns  <- PRICE  ~ NROOM + NBATH + PATIO
+ FIREPL + AC + GAR + AGE +
+                       LOTSZ + SQFT
R> mod_ns   <- regimes(formula = form_ns, data = baltim,
+                      rgv = ~ CITCOU)
R> summary(mod_ns)

                --------------------------------
                          Regimes Model
                --------------------------------

Call:
regimes(formula = form_ns, data = baltim, rgv = ~CITCOU)


Coefficients:
             Estimate Std. Error z-value  Pr(>|z|)
Intercept_0  8.150793   6.900731  1.1811 0.2375435
NROOM_0      1.263246   1.613613  0.7829 0.4337046
NBATH_0      4.379132   2.634815  1.6620 0.0965075 .
PATIO_0     11.084383   5.722933  1.9368 0.0527654 .
FIREPL_0     7.466420   4.270493  1.7484 0.0803993 .
AC_0        12.572501   4.952575  2.5386 0.0111304 *
GAR_0        0.106564   3.101820  0.0344 0.9725939
AGE_0        0.048081   0.071326  0.6741 0.5002429
LOTSZ_0      0.162311   0.046495  3.4909 0.0004814 ***
SQFT_0      -0.161628   0.266790 -0.6058 0.5446316
Intercept_1 12.650330   7.774536  1.6271 0.1037054
NROOM_1      1.880380   1.676226  1.1218 0.2619502
NBATH_1     13.672410   2.644772  5.1696 2.346e-07 ***
PATIO_1      6.822112   3.246130  2.1016 0.0355871 *
FIREPL_1    11.439887   3.079421  3.7149 0.0002032 ***
AC_1         1.948290   3.072815  0.6340 0.5260542
GAR_1        9.068525   2.250954  4.0287 5.607e-05 ***
AGE_1       -0.226210   0.124555 -1.8161 0.0693479 .
LOTSZ_1      0.047377   0.017334  2.7332 0.0062720 **
SQFT_1       0.139858   0.227119  0.6158 0.5380311
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the coefficients table printed by the `summary` method, the different regimes are indicated with numbers inherited from the regime variable.

Since this basic specification does not account explicitly for space, one can obtain the spatial LM tests for spatial dependence by estimating two separate equations, and then using the function `lm.LMtests` available from the package **spdep**:

```
R> eq1 <- lm(form_ns, data = subset(baltim, CITCOU == 0))
R> eq2 <- lm(form_ns, data = subset(baltim, CITCOU == 1))
R> library("spdep")
R> st1 <- lm.LMtests(eq1, subset(wlis, baltim$CITCOU == 0),
test = "all")
R> st2 <- lm.LMtests(eq2, subset(wlis, baltim$CITCOU == 1),
 test = "all")
```

Table 1 LM tests results for the regime model with two equations using a spatial weight matrix based on the queen contiguity criterion

	First equation		Second equation	
	Statistic	*p* value	Statistic	*p* value
LMerr	0.0062	0.9371	7.1964	0.0073
LMlag	0.0103	0.9190	30.0204	0.0000
RLMerr	0.0451	0.8318	1.0816	0.2983
RLMlag	0.0492	0.8244	23.9056	0.0000
SARMA	0.0554	0.9727	31.1020	0.0000

Table 1 reports the results of the five tests implemented in `lm.LMtests`.[10] While none of the tests is statistically significant in the first equation, the equation for the second regime points at a spatial lag specification.

Interestingly, it is also possible to test various types of restrictions. As an example, we can consider restrictions on the coefficients for the same variable in different regimes. The code below shows how to implement those tests for the variable `NBATH`: $H_0 : \beta_{\texttt{NBATH\_0}} = \beta_{\texttt{NBATH\_1}}$. There are multiple ways to test linear hypothesis in R. We choose the implementation provided by the function `linearHypothesis` from the **car** package (Fox and Weisberg 2019).

```
R> library("car")
R> linearHypothesis(mod_ns[[1]], c("NBATH_0 - NBATH_1=0"))

Linear hypothesis test

Hypothesis:
NBATH_0 - NBATH_1 = 0

Model 1: restricted model
Model 2: PRICE ~ Intercept_0 + NROOM_0 + NBATH_0 + PATIO_0
 + FIREPL_0 +
   AC_0 + GAR_0 + AGE_0 + LOTSZ_0 + SQFT_0
+ Intercept_1 + NROOM_1 +
   NBATH_1 + PATIO_1 + FIREPL_1 + AC_1 + GAR_1
+ AGE_1 + LOTSZ_1 +
   SQFT_1 - 1

 Res.Df   RSS Df Sum of Sq      F  Pr(>F)
1    192 31520
2    191 30530  1    990.51 6.1968 0.01365 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The result shows that we can reject the null hypothesis that `NBATH` has the same effect on housing price regardless of whether the dwelling is in Baltimore County or another county.

One can also perform a Wald test for the joint significance of the coefficients using the function `wald.test` from the library `aods3` (Lesnoff and Lancelot 2022). In our

[10] This table was produced with the package **xtable** (Dahl et al. 2019).

example below, the null hypothesis is

$$H_0 : \beta_{\texttt{AGE\_0}} = \beta_{\texttt{AGE\_1}},$$
$$\beta_{\texttt{LOTSZ\_0}} = \beta_{\texttt{LOTSZ\_1}},$$
$$\beta_{\texttt{SQFT\_0}} = \beta_{\texttt{SQFT\_1}}.$$

```
R> library("aods3")
R> first <- c(0,0,0,0,0,0,0,1,0,0,
+             0,0,0,0,0,0,0,-1,0,0)
R> second <- c(0,0,0,0,0,0,0,0,1,0,
+              0,0,0,0,0,0,0,0,-1,0)
R> third <- c(0,0,0,0,0,0,0,0,0,1,
+             0,0,0,0,0,0,0,0,0,-1)
R> lmat <- rbind(first, second, third)
R> wald.test(coef(mod_ns[[1]]), varb = vcov(mod_ns[[1]]), L = lmat)

Chi-squared test:
X2 = 8.998, df = 3, P(> X2) = 0.02932
```

In the example below, we show that it is possible to identify regimes using a (clustering) data driven procedure. For an illustrative purpose, we use the (scaled) geographical coordinates of the dwellings to identify two regimes using the `kmeans` function. The results from this model are different from the previous one.

```
R> df <- scale(baltim[,c("X", "Y")])
R> vrreg <- kmeans(df, centers = 2)
R> mod_ns3 <- regimes(formula = form_ns,
+                    data = baltim, rgv = ~ vrreg$cluster,
+                    vc = "groupwise")
R> summary(mod_ns3)

                --------------------------------
                          Regimes Model
                --------------------------------

Call:
regimes(formula = form_ns, data = baltim, rgv = ~vrreg$cluster,
   vc = "groupwise")


Coefficients:
             Estimate Std. Error z-value  Pr(>|z|)
Intercept_1 19.207028   7.891701  2.4338  0.014940 *
NROOM_1      3.040837   1.726507  1.7613  0.078193 .
NBATH_1      9.992617   3.215458  3.1077  0.001886 **
PATIO_1      8.520687   3.796738  2.2442  0.024819 *
FIREPL_1     5.455361   4.236263  1.2878  0.197824
AC_1        -2.661242   3.590460 -0.7412  0.458573
GAR_1       13.295143   2.477742  5.3658 8.058e-08 ***
AGE_1       -0.595531   0.109828 -5.4224 5.881e-08 ***
LOTSZ_1      0.042853   0.021524  1.9910  0.046481 *
SQFT_1       0.190504   0.273335  0.6970  0.485826
Intercept_2 16.828253   7.096335  2.3714  0.017721 *
NROOM_2     -0.140388   1.628934 -0.0862  0.931320
NBATH_2      6.205570   2.531654  2.4512  0.014238 *
```

```
PATIO_2      9.785026   4.839851  2.0218  0.043201 *
FIREPL_2    10.432588   3.504153  2.9772  0.002909 **
AC_2        11.741856   3.888001  3.0200  0.002528 **
GAR_2       -1.686457   3.247405 -0.5193  0.603534
AGE_2       -0.055655   0.069501 -0.8008  0.423256
LOTSZ_2      0.134456   0.023130  5.8129 6.139e-09 ***
SQFT_2       0.123873   0.247989  0.4995  0.617420
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4 Endogenous variables and the function ivregimes

4.1 Endogenous variables

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

$$y = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + X\beta + \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + Y\pi + \varepsilon, \tag{2}$$

where the difference with Equation (1) is given by the presence of the $n_1 \times q$ matrix Y_1, the $n_2 \times q$ matrix Y_2 and the $n \times r$ matrix Y, with the corresponding vectors of parameters π_1, π_2 and π. Since those three matrices contain endogenous variables, the model is estimated using IV techniques.

4.2 The function ivregimes

The function `ivregimes` has four arguments: `formula`, `data`, `rgv` and `vc`. The right-hand side of the `formula` has four parts. The first part must contain all the regressors (exogenous and endogenous) that do not vary by regimes. The second part has all the regressors (exogenous and endogenous) that vary by regimes. The third part includes all the exogenous regressors and external instruments that do not vary by regimes. The fourth part has all the exogenous regressors and external instruments that vary by regimes. Let H be the matrix of instruments (exogenous regressors and additional instruments for the endogenous variables) for the endogenous variables. Then the `formula` for `ivregimes` has the following structure:

```
y ~ fixed Xs | varying Xs | fixed Hs | varying Hs
```

The following `formula` states that none of the regressors (exogenous and endogenous) is fixed (note the `0`), and they all vary by regime. The instrument matrix is made up of the exogenous variables MA90, PS90, and RD90, and the external instruments FH90, FP89, and GI89. The function `ivregimes` checks internally that the instruments are at least as many as the endogenous variables.

```
R> form_nse   <- HR90  ~  0 | MA90 + PS90 + RD90 +
+                         UE90 | 0 | MA90 + PS90 + RD90 +
+                         FH90 + FP89 + GI89
```

The argument `vc` determines how the variance-covariance matrix should be estimated. Specifically, it takes on three values: `"homoskedastic"`, `"robust"` and `"OGMM"`.,[11]

We use `ivregimes` to estimate the previous model, `form_nse`, and we set `vc = "robust"`:

```
R> mod_nse <- ivregimes(formula = form_nse, data = natreg,
+                       rgv = ~ REGIONS, vc = "robust")
R> summary(mod_nse)
```

```
               ------------------------------------
                          IV Regimes Model
               ------------------------------------

Call:
ivregimes(formula = form_nse, data = natreg, rgv = ~REGIONS,
   vc = "robust")


Coefficients:
               Estimate Std. Error z-value  Pr(>|z|)
(Intercept)_2  -3.444691   2.396091 -1.4376  0.150539
MA90_2          0.011485   0.048761  0.2355  0.813794
PS90_2          0.740137   0.281651  2.6278  0.008593 **
RD90_2          0.130083   0.597848  0.2176  0.827752
UE90_2          1.004032   0.222674  4.5090 6.514e-06 ***
(Intercept)_0 -11.130505   5.568414 -1.9989  0.045623 *
MA90_0          0.179010   0.088457  2.0237  0.043002 *
PS90_0          0.544240   0.373615  1.4567  0.145203
RD90_0         -1.142647   1.185927 -0.9635  0.335294
UE90_0          1.414407   0.450072  3.1426  0.001674 **
(Intercept)_1  15.613820   2.064515  7.5629 3.930e-14 ***
MA90_1         -0.023799   0.050756 -0.4689  0.639146
PS90_1          2.261376   0.335955  6.7312 1.683e-11 ***
RD90_1          5.965739   0.452688 13.1785 < 2.2e-16 ***
UE90_1         -1.211717   0.178151 -6.8016 1.034e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Endogenous variables:
UE90_2  UE90_0  UE90_1
Instruments:
FH90_2 FP89_2 GI89_2 FH90_0 FP89_0 GI89_0 FH90_1 FP89_1 GI89_1
```

[11] When `vc = "OGMM"` a two step procedure is adopted. In the first step, the model is estimated by two stage least squares using the matrix of instruments which is made up of all the exogenous variables and the external instruments. In the second step, the optimal weighted GMM is obtained by using the residuals from the first step to estimate the weighting matrix for the moments conditions (see Anselin and Rey 2014, for additional details).

5 The spatial models and the function spregimes

5.1 The spatial model

A general spatial model is one that contains spatial lag of the dependent variable, spatial lag of the error term, and spatial lag of (some of) the regressors. This is combined with the fact that **hspm** allows for additional endogenous variables and regimes. For this reason, we decided to present each model separately. It is worth emphasizing again that our presentation of the function is not intended to guide users' choice in terms of model specification, but rather to illustrate the arguments of the function. The general model is estimated following a series of steps that alternate IV with GMM techniques. These steps are an adaptation of the general cross-sectional model in Kelejian and Prucha (2010a) and Arraiz et al. (2010) to spatial regimes models.[12]

5.2 The function spregimes

`spregimes` is used to estimate the general model as well as all of the nested specifications that derive from it. The function has eleven arguments. In this section we describe the `formula`, and we delay the discussion of the other arguments to the next sections. In `spregimes`, the right-hand side of `formula` must be specified with six parts. Specifically, the `formula` for `spregimes` has the following structure:

```
y ~ fixed Xs | varying Xs | WXs | fixed Hs |
 varying Hs | W external
instruments
```

Since the specification of the formula is the trickiest part, we use three examples.

`form_sp_b` below is based on the Baltimore data. The variables AC, AGE and NROOM are the regressors that do not vary by regimes, while PATIO, FIREPL, and SQFT are those that vary. The third part is used to specify the spatially weighted regressors (in this case, AGE, NROOM and NBATH). It is important to stress that the spatial lag of one regressor varies only if the regressor itself vary. Vice-versa, if the regressor is fixed, also the lag would be so. For example, since AGE and NROOM vary by regimes also their lags vary. On the other hand, since NBATH is fixed, also the lag of NBATH will not vary. The next three parts of the formula serve to specify the fixed instruments (part four), the instruments that vary (part five), and the spatial lag of the external instruments (part six). Since there are no endogenous variables in Baltimore data, part four and part five of the `formula` are the same as part one and part two. The sixth part is set to `0` indicating that there are no external instruments to be lagged.

```
R> form_sp_b <- PRICE ~AC+AGE +NROOM + PATIO + FIREPL + SQFT |
+                       NBATH + GAR + LOTSZ - 1 |
+                       AGE + NROOM + NBATH |
+                       AC + AGE + NROOM + PATIO+FIREPL+SQFT |
+                       NBATH + GAR + LOTSZ - 1 | 0
```

[12] For implementation details in the context of cross-sectional models see also Bivand and Piras (2015). For additional information it is also possible to consult the details of the help function.

The second and third formulas are specified in terms of `natreg` data. The formula `form_sp_n` should be interpreted in the following way. The regressor MA90 is fixed. The intercept, PS90, RD90, and UE90 are the regressors that vary by regimes. The spatial lag of MA90 is also considered among the regressors. Since MA90 is fixed, also its spatial lag is fixed. Next, we have one instrument fixed (MA90), and five instruments that change by regimes, namely PS90, RD90, FH90, FP89, and GI89. None of the additional instruments is spatially lagged (the `0` in the last line below).

```
R> form_sp_n <-  HR90   ~ MA90 -1 |
+                PS90 + RD90 + UE90 |
+                MA90 |
+                MA90 -1 |
+                PS90 + RD90 + FH90 + FP89 + GI89 | 0
```

In `form_sp_n2` all of the regressors vary by regime (since the first part of the formula is `0`). The spatial lag of MA90 is also included. Since MA90 varies, so does its spatial lag. Furthermore, all of the instruments vary by regimes, and none of the external instruments is lagged.

```
R> form_sp_n2 <-  HR90   ~   0 |
+                 MA90 +  PS90 + RD90 + UE90 |
+                 MA90 |
+                 0 |
+                 MA90 + PS90 + RD90 + FH90 + FP89 + GI89 | 0
```

Linear model with regimes, spatially lagged regressors and potential endogeneity

The first case that we consider is that of a linear model with regimes that includes spatial lag of the regressors (both exogenous and endogenous):

$$\begin{aligned} y = &\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + X\beta + \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + Y\pi + \\ &W \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + WX\delta + W \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + WY\theta + \varepsilon. \end{aligned} \tag{3}$$

Compared to Equation (2), Equation (3) includes the spatial lags of the exogenous and endogenous variables whether or not they change by regimes, and the relative vectors of parameters. Interestingly, there is no limitation as for the specification of the spatial weighting matrix. This means that the spatial weighting matrix does not necessarely need to be block-diagonal, but it can have a structure where observations that are in different regimes are considered neighbours.[13]

[13] The specification of the spatial weighting matrix was one of the crucial aspects in the implementation of the package. It is quite reasonable that W is not block-diagonal when one is constructing the spatial lag for variables that are fixed. At the same time, a spatial weighting matrix that allows for interactions across regimes can be difficult to justify for variables that vary across regimes. However, also the opposite situation is true: forcing the spatial weighting matrix to a block-diagonal structure, while it seems a logical choice if the regressors are different by regimes, it would be a limitation for variables that are fixed by regimes. Because of this, in the end we decided to leave the decision to the users. Additionally, this was one of the reasons for not implementing the impacts for the regimes models. We delay this discussion to Sect. 6.

For models with endogenous variables, we use the specification reflected in `form_sp_n` based on the `natreg` data. The arguments `model = "ols"` and the regimes variables is set to `~ REGION`.[14] The spatial weighting matrix is `listw = w_6`. The argument `listw`, as in other spatial packages, can be of class `listw`, or `matrix`, or `Matrix` (Bates et al. 2022).

```
R> mod_s_b_olse
<- spregimes(formula = form_sp_n, data = natreg,
+
 rgv = ~ REGIONS, listw = ws_6, model = "ols")
R> summary(mod_s_b_olse)

------------------------------------------------------------
               Regimes Model with spatially lagged regressors
                    and additional endogenous variables
------------------------------------------------------------

Call:
spregimes(formula = form_sp_n, data = natreg, model = "ols",
   listw = ws_6, rgv = ~REGIONS)


Coefficients:
              Estimate Std. Error  z-value  Pr(>|z|)
MA90           0.036269   0.035040   1.0351  0.300640
(Intercept)_2 -2.585496   2.862794  -0.9031  0.366453
PS90_2         0.732342   0.249492   2.9353  0.003332 **
RD90_2         0.138214   0.750891   0.1841  0.853962
UE90_2         1.009683   0.233377   4.3264 1.516e-05 ***
(Intercept)_0 -2.285820   3.424251  -0.6675  0.504428
PS90_0         0.483703   0.238178   2.0309  0.042270 *
RD90_0        -0.402834   0.993990  -0.4053  0.685279
UE90_0         1.110973   0.352428   3.1523  0.001620 **
(Intercept)_1 15.160386   1.832372   8.2736 2.220e-16 ***
PS90_1         2.319289   0.180390  12.8571 < 2.2e-16 ***
RD90_1         5.958644   0.277295  21.4885 < 2.2e-16 ***
UE90_1        -1.195927   0.114842 -10.4136 < 2.2e-16 ***
W_MA90        -0.050150   0.055819  -0.8984  0.368948
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Endogenous variables:
UE90_2 UE90_0 UE90_1
Instruments:
X WX FH90_2 FP89_2 GI89_2 FH90_0 FP89_0 GI89_0 FH90_
1 FP89_1 GI89_1
```

The `summary` method prints a description reflecting the fact that the model contains endogenous variables. In the bottom part of the output, a list of the endogenous variables and the instruments is given. The function `spregimes` checks internally that the instruments are at least as many as the endogenous variables.

[14] It might seems counterintuitive that the `model` argument is set to `"ols"` when there are endogenous variables in the model. However, this is consistent with other spatial libraries, such as **sphet**.

Spatial Lag (and Durbin) regimes model

In this section we include both the spatial lag and the spatial Durbin model (with or without additional endogenous variables). Regimes models that include the spatial lag of the dependent variable can be specified in two different ways depending on whether the spatial lag coefficient is allowed to vary by regimes. When the coefficient of the spatial lag is not allowed to change, the model can be written in the following way:

$$\begin{aligned} y = \lambda W y &+ \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + X\beta + \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix}\begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + Y\pi \\ &+ W\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + WX\delta + W\begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + WY\theta + \varepsilon \end{aligned} \tag{4}$$

where λ is a scalar parameter. On the contrary, when the coefficient of the spatial lag is allowed to vary, the model can be written as[15]:

$$\begin{aligned} y = W\begin{bmatrix} y_1 & 0 \\ 0 & y_2 \end{bmatrix}\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} &+ \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + X\beta + \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix}\begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + Y\pi \\ &+ W\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + WX\delta + W\begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + WY\theta + \varepsilon \end{aligned} \tag{5}$$

No endogenous variables

In the example below, we are assuming that the spatial process is different by regimes (`wy_rg = TRUE`), and that the `model = "lag"`.[16]

```
R> mod_s_b_lag <- spregimes(formula = form_sp_b, data = baltim,
+                    rgv = ~ CITCOU, listw = wlis, model = "lag",
+                    wy_rg = TRUE)
R> summary(mod_s_b_lag)
```

```
------------------------------------------------------------
                  Spatial Durbin Regimes Model
------------------------------------------------------------

Call:
spregimes(formula = form_sp_b, data = baltim, model = "lag",
   listw = wlis, wy_rg = TRUE, rgv = ~CITCOU)


Coefficients:
```

[15] Also in this case no restriction on W is imposed (i.e., W is not necessarily block diagonal). For the spatial model in Eq. (5), this means that observations that equal zero in y_1 (because they belong to a different regimes) may not necessarily be zero when one takes the spatial lag of the first column of $\begin{bmatrix} y_1 & 0 \\ 0 & y_2 \end{bmatrix}$. Clearly, the same is also true if we consider y_2 and the spatial lag of the second column of $\begin{bmatrix} y_1 & 0 \\ 0 & y_2 \end{bmatrix}$.

[16] Note that we do not allow explicitly the argument `model` to be set to Durbin. This is due to the fact that **hspm** deals with spatially lagged regressors through the `formula` argument.

```
              Estimate Std. Error z-value  Pr(>|z|)
(Intercept)   9.697585  11.003865  0.8813  0.378161
AC            5.491020   2.489142  2.2060  0.027385 *
AGE          -0.022360   0.064788 -0.3451  0.729993
NROOM         1.537288   1.158262  1.3272  0.184430
PATIO         4.565648   2.795709  1.6331  0.102450
FIREPL        6.756735   2.465632  2.7404  0.006137 **
SQFT         -0.014176   0.168549 -0.0841  0.932972
NBATH_0       3.465790   2.084328  1.6628  0.096355 .
GAR_0         1.483856   2.937730  0.5051  0.613487
LOTSZ_0       0.116911   0.043997  2.6573  0.007878 **
NBATH_1      12.940427   2.207887  5.8610 4.601e-09 ***
GAR_1         6.219459   2.098825  2.9633  0.003044 **
LOTSZ_1       0.033165   0.016492  2.0109  0.044333 *
W_AGE         0.038113   0.116357  0.3276  0.743249
W_NROOM      -2.562542   2.378078 -1.0776  0.281226
W_NBATH_0    -2.013042   5.530003 -0.3640  0.715842
W_NBATH_1   -14.556912   5.593529 -2.6025  0.009256 **
W_PRICE_0     0.533085   0.268570  1.9849  0.047155 *
W_PRICE_1     0.793301   0.140087  5.6629 1.488e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Endogenous variables:
W_PRICE_0 W_PRICE_1
Instruments:
X WX WWX WWWX
```

Note that since the `wy_rg` is `TRUE` and there are two regimes, the function estimates two coefficients for the spatially lagged dependent variable (`W_PRICE_0` and `W_PRICE_1`).

With endogenous variables

The model specification below does not allow for a varying λ (`wy_rg = FALSE`) but allows for heteroskedasticity in the error term (`het = TRUE`). When `het` is set to `TRUE`, a robust estimator of the variance-covariance is calculated.

```
R> mod_s_n_lag <- spregimes(formula = form_sp_n, data = natreg,
+                       rgv = ~REGIONS, listw = ws_6, model = "lag",
+                       het = TRUE, wy_rg = FALSE)
R> summary(mod_s_n_lag)

 ------------------------------------------------------------
                        Spatial Durbin Regimes Model
                   with additional endogenous variables
 ------------------------------------------------------------


Call:
spregimes(formula = form_sp_n, data = natreg, model = "lag",
   listw = ws_6, wy_rg = FALSE, rgv = ~REGIONS, het = TRUE)


Coefficients:
               Estimate Std. Error z-value  Pr(>|z|)
MA90           0.026297   0.036287  0.7247  0.468631
```

```
(Intercept)_2  6.117144   2.064886  2.9625  0.003052 **
PS90_2         1.457928   0.213599  6.8255 8.760e-12 ***
RD90_2         2.897389   0.446956  6.4825 9.022e-11 ***
UE90_2         0.089726   0.108346  0.8281  0.407587
(Intercept)_0  4.303163   2.833462  1.5187  0.128839
PS90_0         0.615684   0.347250  1.7730  0.076224 .
RD90_0         1.149395   0.788002  1.4586  0.144670
UE90_0         0.499441   0.276218  1.8081  0.070585 .
(Intercept)_1 16.551061   2.324691  7.1197 1.082e-12 ***
PS90_1         2.228211   0.362886  6.1403 8.239e-10 ***
RD90_1         5.732077   0.549753 10.4266 < 2.2e-16 ***
UE90_1        -1.163982   0.187167 -6.2189 5.005e-10 ***
W_MA90        -0.099521   0.057046 -1.7446  0.081057 .
W_HR90         0.055918   0.075654  0.7391  0.459828
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Endogenous variables:
W_HR90 UE90_2 UE90_0 UE90_1
Instruments:
X WX WWX WWWX FH90_2 FP89_2 GI89_2 FH90_0
FP89_0 GI89_0 FH90_1 FP89_1 GI89_1
```

The last row in the coefficients table reports the lag of the homicides rate. The other endogenous variable are UE90_2, UE90_0, and UE90_1. As a consequence, the external instruments are also different by regimes.

Spatial error regimes model

The spatial error regimes model is slightly different from the previous specification. In fact, the spatial error coefficient can be different by regime if and only if all the explanatory variables in the model vary by regimes, that is[17]:

$$
\begin{aligned}
y = &\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix}\begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} \\
&+ W\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + W\begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix},
\end{aligned} \tag{6}
$$

where

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

where ρ_1 and ρ_2 are the spatial error parameters for the first and the second regime, respectively. Alternatively, the hybrid model can include a spatial error process that

[17] The reason for this is related to the peculiar estimation of the error model that involves the so-called spatial Cochrane-Orcutt transformation (Kelejian and Prucha 1999, 2010b). Further details can be obtained by writing to the authors.

does not vary by regimes, such as:

$$
\begin{aligned}
y = &\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + X\beta + \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix}\begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + Y\pi \\
&+ W\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + WX\delta + W\begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + WY\theta + \varepsilon,
\end{aligned} \tag{7}
$$

and

$$\varepsilon = \rho W \varepsilon + u,$$

where the spatial error coefficient ρ is a scalar parameter.

The spatial error regimes model is obtained from the `natreg` data setting the argument `model = "error"`. For the following example we use the formula `formula = form_sp_n2`, where all the exogenous and endogenous variables are different by regime. We also set `weps_rg = TRUE`, and we allow for heteroskedasticity by setting `het = TRUE`:

```
R> mod_s_n_error <- spregimes(formula=form_sp_n2, data = natreg,
+  rgv = ~ REGIONS, listw = ws_6, model = "error",
+ weps_rg = TRUE, het = TRUE)
R> summary(mod_s_n_error)
```

```
             --------------------------------------------
                     Spatial Error Regimes Model
                  with spatially lagged regressors
                 and additional endogenous variables
            --------------------------------------------


Call:
spregimes(formula = form_sp_n2, data = natreg, model = "error",
   listw = ws_6, weps_rg = TRUE, rgv = ~REGIONS, het = TRUE)


Coefficients:
              Estimate Std. Error z-value  Pr(>|z|)
(Intercept)_2 -0.506749   2.305999 -0.2198 0.8260639
MA90_2         0.047852   0.048794  0.9807 0.3267354
PS90_2         1.014640   0.279003  3.6367 0.0002762 ***
RD90_2         0.832787   0.514065  1.6200 0.1052315
UE90_2         0.878607   0.192386  4.5669 4.950e-06 ***
W_MA90_2      -0.089340   0.015470 -5.7750 7.695e-09 ***
We_2           0.413651   0.058780  7.0372 1.961e-12 ***
(Intercept)_0 -5.590587   4.800468 -1.1646 0.2441841
MA90_0         0.227225   0.090092  2.5222 0.0116636 *
PS90_0         0.648210   0.365976  1.7712 0.0765302 .
RD90_0        -0.805318   1.143382 -0.7043 0.4812272
UE90_0         1.304638   0.440551  2.9614 0.0030627 **
W_MA90_0      -0.188891   0.055211 -3.4213 0.0006233 ***
We_0           0.110381   0.131750  0.8378 0.4021372
(Intercept)_1 16.311945   2.390988  6.8223 8.962e-12 ***
MA90_1        -0.005598   0.054441 -0.1028 0.9181016
PS90_1         2.161770   0.360342  5.9992 1.983e-09 ***
RD90_1         5.861047   0.512513 11.4359 < 2.2e-16 ***
```

```
UE90_1        -1.195113   0.207373 -5.7631 8.258e-09 ***
W_MA90_1      -0.036756   0.016538 -2.2225 0.0262507 *
We_1           0.236253   0.052229  4.5234 6.085e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Endogenous variables:
UE90_2 UE90_0 UE90_1
Instruments:
X WX FH90_2 FP89_2 GI89_2 FH90_0 FP89_0 GI89_0 FH90_1 FP89_1 GI89_1
```

In this case we have the endogenous variable `UE90` that varies by regimes, and the instruments matrix that includes all the exogenous variables in the model and the external instruments that are also different by regimes.

Spatial SARAR regimes model

The spatial SARAR regimes model is selected by the argument `model = "sarar"`. One can then choose to set `wy_rg` and `weps_rg` in order to have four possible combinations. However, as for the error regimes model, `weps_rg` can be `TRUE` only if all variables are different by regimes. Since this model is just a combinations of the arguments presented in the previous sections, to save space we do not include output for the SARAR model.

6 Impacts

As noted by LeSage and Pace (2009), models that contain spatial lag of the dependent variable needs to be correctly interpreted. This means that appropriate impacts measures have to be calculated. LeSage and Pace (2009) suggested the computation of three average spillover impacts for spatial models including a spatial lag of the dependent variable. These spillover impacts defined on the j-th variable of a spatial lag model (without regimes) are:

$$ATI_j = \beta_j n^{-1} e'(I_n - \lambda W)^{-1} e, \tag{8}$$

$$ADI_j = \beta_j n^{-1} tr[(I_n - \lambda W)^{-1}], \tag{9}$$

and

$$AII_j = ATI_j - ADI_j, \tag{10}$$

where e is a vector of ones, I_n is a diagonal matrix whose diagonal elements are one, and tr indicates the trace operator.

For a variety of reasons, the impacts measures suggested by LeSage and Pace (2009) may not be straightforwardly extended to spatial regimes models. This is mainly, but not exclusively related to how the spatial weighting matrix can be defined in this context. Of course, a general treatment of impacts measures for spatial regimes models is outside of the scope of the present paper. However, in what follows, we show a few (simple) situations in which the impacts can be calculated using matrix algebra.

The first and easiest option relates to the case where all coefficients, including the spatial lag of the dependent variable, vary by regimes. Interestingly, this scenario corresponds to the estimation of two separate equations. In R this can be achieved using the function `spreg` from the **sphet** package (Bivand and Piras 2015; Bivand et al. 2021; Piras and Postiglione 2022). Basically, one can take advantage of the function `subset` that can be applied both to the data and the `listw` object `wlis`:

```
R> library("sphet")
R> ff <- PRICE  ~  NROOM + NBATH + PATIO +
+                  FIREPL + AC + GAR + AGE + LOTSZ + SQFT
R> mod_sphet_1 <- spreg(ff, data = subset(baltim, CITCOU == 0),
+                       listw = subset(wlis, baltim$CITCOU == 0),
+                       model = "lag", het = TRUE)
R> mod_sphet_2 <- spreg(ff, data = subset(baltim, CITCOU == 1),
+                       listw = subset(wlis, baltim$CITCOU == 1),
+                       model = "lag", het = TRUE)
```

Since **sphet** has functions to calculate the effects, one can take advantage of the infrastructure already available in R. For reason of space, the code below illustrates how to obtain impacts and inference only for the first equation. Particularly, the inference for the impacts is obtained using an analytical formula derived in Kelejian and Piras (2020).

```
R> impacts(mod_sphet_1, KPformula = TRUE)

Impact Measures (lag, KP_formula):
            Direct    Indirect       Total
NROOM   1.49562469  0.34661171  1.84223640
NBATH   4.62449820  1.07172959  5.69622780
PATIO  11.59566431  2.68730056 14.28296487
FIREPL  6.68434971  1.54910113  8.23345084
AC     12.04494900  2.79142249 14.83637149
GAR     0.17413512  0.04035590  0.21449102
AGE     0.06355598  0.01472913  0.07828511
LOTSZ   0.14844814  0.03440292  0.18285106
SQFT   -0.19521019 -0.04524005 -0.24045024
========================================================
Results based on Kelejian and Piras formula:
========================================================
Analytical standard errors
           Direct   Indirect      Total
NROOM  2.06242519 0.55236543 2.43939027
NBATH  2.17685545 1.35656560 2.62946178
PATIO  5.56322677 3.88683051 7.98088853
FIREPL 4.10077137 2.50061374 5.98030568
AC     3.04743323 3.58850525 4.56464593
GAR    2.89309374 0.67809572 3.56949626
AGE    0.09365007 0.03326068 0.12263245
```

```
LOTSZ  0.04745690 0.04484869 0.06705187
SQFT   0.31475415 0.07958076 0.37440748

Analytical z-values:
             Direct    Indirect      Total
NROOM   0.72517767  0.62750435  0.7552036
NBATH   2.12439379  0.79003153  2.1663094
PATIO   2.08434148  0.69138609  1.7896460
FIREPL  1.63002252  0.61948837  1.3767609
AC      3.95248987  0.77787889  3.2502787
GAR     0.06018993  0.05951357  0.0600900
AGE     0.67865386  0.44283906  0.6383719
LOTSZ   3.12806260  0.76708874  2.7270092
SQFT   -0.62019893 -0.56847979 -0.6422154

Analytical p-values:
      Direct     Indirect Total
NROOM  0.4683430  0.53033  0.4501268
NBATH  0.0336372  0.42951  0.0302876
PATIO  0.0371291  0.48932  0.0735108
FIREPL 0.1030967  0.53559  0.1685862
AC     7.7342e-05 0.43664  0.0011529
GAR    0.9520044  0.95254  0.9520840
AGE    0.4973572  0.65788  0.5232316
LOTSZ  0.0017596  0.44303  0.0063911
SQFT   0.5351268  0.56971  0.5207334
```

The `summary` method for the impacts reports the direct, indirect, and total impact for each variables, along with the standard error, a z-value and a p-value to determine the statistical significance of the impacts.

As an alternative, one can calculate the impacts manually taking advantage of sparse matrix representations from the package **Matrix**. In doing this we will use the model described by the `formula = form` which is written in such a way that all variables are different by regimes, and there are no spatially weighted regressors:

```
R> form <- PRICE  ~  -1 | NROOM + NBATH + PATIO +
+    FIREPL + AC + GAR + AGE +
+    LOTSZ + SQFT | 0 | -1 | NROOM +
+    NBATH + PATIO + FIREPL +
+    AC + GAR + AGE + LOTSZ + SQFT | 0
```

In `spregimes` we set the argument `wy_rg` to `FALSE`:

```
R> mod_nwy <- spregimes(formula = form,
+                data = baltim,
+                rgv = ~ CITCOU, model = "lag",
+                wy_rg = FALSE, listw = wlis, het = TRUE)
```

To calculate the impact, we follow a series of steps described in what follows. First of all, we create a sparse spatial weighting matrix with the function `listw2dgCMatrix`, we separate the betas from the spatial lag, and calculate the inverse term of $(I_n - \lambda W)$:

```
R> library(Matrix)
R> Ws       <- listw2dgCMatrix(wlis)
R> n        <- dim(Ws)[1]
R> I        <- Diagonal(n)
R> cf       <- length(coefficients(mod_nwy))
R> betas    <- coefficients(mod_nwy)[-cf]
R> k1       <- length(betas)
R> betas    <- betas[-c(1,((k1/2)+1))]
R> k        <- length(betas)
R> Wprice   <- coefficients(mod_nwy)[cf]
R> IlW      <- I - Wprice * Ws
R> IlWi     <- solve(IlW)
```

Since we have multiple betas, we use a simple loop to iterate over and obtain the three impacts.[18]

```
R> ATI      <- vector(mode = "numeric", length = k)
R> for(i in 1:k) ATI[i] <- sum(IlWi * betas[i])/n
R> ADI      <- vector(mode = "numeric", length = k)
R> for(i in 1:k) ADI[i] <- sum(diag(IlWi * betas[i]))/n
R> AII      <- ATI - ADI
R> effects <- cbind(ADI, AII, ATI)
R> rn       <- rownames(coefficients(mod_nwy))
R> rownames(effects) <- rn[-c(1,((k1/2)+1), cf)]
R> colnames(effects)
<- c("Average Direct", "Average Indirect", "Average Total")
R> effects
```

```
        Average Direct Average Indirect Average Total
NROOM_0      1.99703015       1.30031781    3.29734796
NBATH_0      4.68741092       3.05209407    7.73950499
PATIO_0     12.98101581       8.45227399   21.43328980
FIREPL_0     3.31133478       2.15609543    5.46743020
AC_0        11.32510082       7.37406506   18.69916587
GAR_0        1.02686125       0.66861583    1.69547707
AGE_0        0.08743152       0.05692892    0.14436043
LOTSZ_0      0.11624634       0.07569099    0.19193733
SQFT_0      -0.14263578      -0.09287383   -0.23550962
NROOM_1      0.97618165       0.63561704    1.61179869
NBATH_1     12.23459939       7.96626302   20.20086241
PATIO_1      3.71166347       2.41675975    6.12842322
FIREPL_1     9.65643706       6.28755507   15.94399213
AC_1         1.67440848       1.09025052    2.76465900
GAR_1        8.07706396       5.25918454   13.33624850
AGE_1       -0.14969278      -0.09746883   -0.24716161
LOTSZ_1      0.03550754       0.02311987    0.05862741
SQFT_1       0.02661675       0.01733085    0.04394761
```

[18] Note that the impacts are different from those in the previous case since now the spatial lag of price is unique.

An alternative way to produce inference would be to generate Monte Carlo samples for the effects from a multivariate normal distribution. For this purpose, we use the package **mvtnorm** (Genz et al. 2021; Genz and Bretz 2009). The function `rmvnorm` generates S samples from a multivariate normal with means equal to the vector of the estimated coefficients, and variance-covariance matrix as the estimated variance-covariance matrix of the model. We `apply` to the generated samples the function `eff` which calculates the three impacts. Finally, we compute the standard deviation of the S samples of impacts.

```
R> set.seed(2512)
R> library("mvtnorm")
R> S   <- 399
R> mvn <- rmvnorm(S, coefficients(mod_nwy), vcov(mod_nwy))
R> eff <- function(x){
+       betas <- x[-c(1,length(x))]
+       IlW <- I - x[length(x)] * Ws
+       IlWi <- solve(IlW)
+       ATE  <- vector(mode = "numeric", length = k)
+       for(i in 1:k) ATE[i] <- sum(IlWi * betas[i])/n
+       ADE  <- vector(mode = "numeric", length = k)
+       for(i in 1:k) ADE[i] <- sum(diag(IlWi * betas[i]))/n
+       AIE <- ATE - ADE
+       return(rbind(ADE, AIE, ATE))
+  }
R> MCeffects <- apply(mvn, 1, eff)
R> sterr     <- matrix(apply(MCeffects, 1, sd), 18, 3, byrow = T)
R> rownames(sterr) <- rownames(effects)
R> colnames(sterr) <- colnames(effects)
R> sterr

        Average Direct Average Indirect Average Total
NROOM_0      2.27296700       1.67983768    3.85097136
NBATH_0      2.47357046       1.96810173    4.19753337
PATIO_0      5.58940086       5.46916593   10.32911754
FIREPL_0     4.84338926       3.78897285    8.44577684
AC_0         2.96899760       3.57652004    5.69314927
GAR_0        3.01809750       2.37546952    5.26623013
AGE_0        0.09421574       0.07561498    0.16605169
LOTSZ_0      0.05014582       0.04551636    0.08840079
SQFT_0       0.39001381       0.28274951    0.66180201
NROOM_1      8.10776574       6.16234288   13.95039547
NBATH_1      1.67090126       1.25988956    2.87049777
PATIO_1      3.26552651       4.07807921    6.58525885
FIREPL_1     3.45158358       2.87005467    6.11982509
AC_1         2.83890691       3.52899772    5.67812563
GAR_1        3.39769311       2.46339063    5.74160747
AGE_1        3.11553987       3.41618382    6.11205622
LOTSZ_1      0.12822484       0.10868092    0.22902261
SQFT_1       0.02540186       0.01736927    0.04099058
```

7 Conclusions

In this paper we described the main functionality of the newly developed package **hspm**. We started from a very simple regime model where space is accounted for by specifying a “spatial” regime variable, and then we presented various spatial specifications containing spatial lag of the dependent variable, spatial lag of the regressors, spatial lag of the error term, and any possible combinations of those lags. Finally, we showed how to manually calculate impacts when the spatial weighting matrix has a block-diagonal structure. Implicitly, we recommend not to use the impacts defined by LeSage and Pace (2009) when the structure of the spatial weighting matrix allows for interactions between observations that belong to different regimes.

However, this is only the beginning stage of the package. In the future, we would like to proceed mainly in two directions. On the one hand, we want to expand the package to include more ways to determine the regimes such as data-driven procedures or endogenous ways of determining the regimes. On the other hand, we intend to explore more the continuous approach where coefficients are allowed to vary smoothly over space.

Funding Open access funding provided by Università degli Studi G. D’Annunzio Chieti Pescara within the CRUI-CARE Agreement.

Appendix A

Checks with other available implementations

In this appendix we compare the results from **hspm** with other available implementations. To the best of our knowledge, the only other implementation of spatial regimes models is available in the Python package **spreg** (Anselin and Rey 2014) which is part of the **PySAL** library. Additionally, we limit our attention to the spatial lag regimes model for a couple of reasons. First of all, the spatial lag model should be fully comparable since it is not based on optimization routines (like the case of the error and SARAR models) that can influence the results. Second, the implementation of **hspm** relies heavily on code available from the **sphet** package as it is the case for the **spreg** package in **PySAL**. Bivand and Piras (2015) compared implementations of spatial cross sectional models using, among others, results from **sphet** and **spreg**. They noticed that comparison for the error model where slightly different.[19] Based on this, we expect the same differences to appear in the regimes specification.

[19] See Table 11 and the discussion on page 22 in Bivand and Piras (2015).

The function to estimate a spatial lag regime model in **spreg** is `spreg.GM_Lag_Regimes`.[20]

We consider two cases: the first case corresponds to a situation where all coefficients (including the spatial parameter) are different by regimes; while the second corresponds to a situation where all coefficients are different by regimes but the spatial process is unique. We use the data in `baltim`, and we consider two spatial weighting matrices: the first was introduced in Sect. 2 and it is based on the queen contiguity criteria. The second matrix is simply a block diagonal version of the queen where observations are neighbors only if they belong to the same regime. The results from **spreg** are reported in Table 2. The first two columns of the table are based on the queen contiguity matrix, with spatial lag coefficient that varies by regimes (column (1)) or is fixed (column (2)). The second two columns are based on the block diagonal version of the matrix, with spatial lag coefficient that vary by regimes (column (3)) or is fixed (column (4)).

Looking at Table 2, it stands clear that columns (1) and (3) are the same. This means that, if the model is specified such that all variables differ by regimes, `spreg.GM_Lag_Regimes` "forces" the spatial weighting matrix to be block-diagonal even if the original spatial matrix is not. The same results can be obtained in R in two different ways. The first way is to consider two separate equations and use the function `spreg` from the **sphet** package. These two equations were estimated in Sect. 6. We set `het = TRUE` to obtain a robust estimator of the variance-covariance matrix in order to match the results for the standard errors in Table 2. Here we report only the `summary` of the two equations.

```
R> summary(mod_sphet_1)

Generalized stsls

Call:
spreg(formula = ff, data = subset(baltim, CITCOU == 0)a,
 listw = subset(wlis,
   baltim$CITCOU == 0), model = "lag", het = TRUE)

Residuals:
    Min.    1st Qu.    Median    3rd Qu.      Max.
-31.01896  -7.47878  -0.33306   5.65189  29.14831

Coefficients:
            Estimate Std. Error t-value  Pr(>|t|)
(Intercept)  1.295358  10.591014  0.1223 0.9026556
NROOM        1.483139   2.054208  0.7220 0.4702941
NBATH        4.585893   2.187158  2.0967 0.0360169 *
PATIO       11.498864   5.502726  2.0897 0.0366477 *
FIREPL       6.628549   4.030935  1.6444 0.1000895
AC          11.944398   3.088166  3.8678 0.0001098 ***
GAR          0.172681   2.868597  0.0602 0.9519986
AGE          0.063025   0.092508  0.6813 0.4956844
LOTSZ        0.147209   0.047680  3.0874 0.0020189 **
SQFT        -0.193581   0.313258 -0.6180 0.5366020
lambda       0.194925   0.216375  0.9009 0.3676599
```

[20] For details on the arguments see the file replication.py available from the additional material associated with the paper.

Table 2 Results obtained with **PySAL** function `spreg.GM_Lag_Regimes`

	Queen		Block diagonal queen	
	(1)	(2)	(3)	(4)
INTERCEPT_0	1.2953	−8.6200	1.2953	−8.7458
	(10.5910)	(10.9209)	(10.5910)	(10.7650)
NROOM_0	1.4831	1.9126	1.4831	1.8052
	(2.0542)	(2.2770)	(2.0542)	(2.3187)
NBATH_0	4.5858	4.5141	4.5858	4.8887
	(2.1871)	(2.4115)	(2.1871)	(2.4599)
PATIO_0	11.4988	12.4829	11.4988	12.1059
	(5.5027)	(5.2289)	(5.5027)	(4.8660)
FIREPL_0	6.6285	3.2656	6.6285	5.4013
	(4.0309)	(4.6409)	(4.0309)	(4.5261)
AC_0	11.9443	10.9414	11.9443	11.0244
	(3.0881)	(3.0495)	(3.0881)	(3.2454)
GAR_0	0.1726	0.9739	0.1726	0.2695
	(2.8685)	(2.8713)	(2.8685)	(2.8965)
AGE_0	0.0630	0.0836	0.0630	0.0849
	(0.0925)	(0.0914)	(0.0925)	(0.0910)
LOTSZ_0	0.1472	0.1128	0.1472	0.1250
	(0.0476)	(0.0487)	(0.0476)	(0.0484)
SQFT_0	−0.1935	−0.1378	−0.1935	−0.2403
	(0.3132)	(0.3812)	(0.3132)	(0.3798)
INTERCEPT_1	1.4189	2.5105	1.4189	2.9856
	(7.4445)	(8.0427)	(7.4445)	(7.4761)
NROOM_1	0.5718	0.9570	0.5718	0.7543
	(1.4873)	(1.5490)	(1.4873)	(1.4755)
NBATH_1	10.3693	11.8217	10.3693	10.8300
	(2.9844)	(3.1891)	(2.9844)	(3.0264)
PATIO_1	1.1846	3.6333	1.1846	1.9709
	(3.1137)	(3.3580)	(3.1137)	(3.1867)
FIREPL_1	7.8769	9.3419	7.8769	8.3739
	(2.5384)	(2.7379)	(2.5384)	(2.5768)
AC_1	0.1223	1.6192	0.1223	0.3770
	(3.1476)	(3.2105)	(3.1476)	(3.1532)
GAR_1	7.9206	7.8052	7.9206	8.0807
	(2.9900)	(3.0377)	(2.9900)	(3.0292)
AGE_1	−0.2087	−0.1456	−0.2087	−0.2112
	(0.1199)	(0.1286)	(0.1199)	(0.1232)

Table 2 continued

	Queen		Block diagonal queen	
	(1)	(2)	(3)	(4)
LOTSZ_1	0.0349	0.0344	0.0349	0.0367
	(0.0249)	(0.0256)	(0.0249)	(0.0248)
SQFT_1	−0.0057	0.0276	−0.0057	0.0145
	(0.2226)	(0.2274)	(0.2226)	(0.2201)
WPRICE_0	0.1949		0.1949	
	(0.2163)		(0.2163)	
WPRICE_1	0.5583		0.5583	
	(0.0949)		(0.0949)	
WPRICE		0.4091		0.4804
		(0.1030)		(0.0930)

Standard errors in parenthesis

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R> summary(mod_sphet_2)
Generalized stsls

Call:
spreg(formula = ff, data = subset(baltim, CITCOU == 1),
 listw = subset(wlis,
   baltim$CITCOU == 1), model = "lag", het = TRUE)

Residuals:
    Min.   1st Qu.    Median   3rd Qu.      Max.
-31.36829  -6.39237  -0.10472   5.51075  58.25803

Coefficients:
             Estimate Std. Error t-value  Pr(>|t|)
(Intercept)  1.4189637  7.4445519  0.1906 0.8488356
NROOM        0.5718325  1.4873796  0.3845 0.7006403
NBATH       10.3693049  2.9844099  3.4745 0.0005118 ***
PATIO        1.1846154  3.1137637  0.3804 0.7036152
FIREPL       7.8769605  2.5384508  3.1031 0.0019153 **
AC           0.1223055  3.1476928  0.0389 0.9690055
GAR          7.9206191  2.9900327  2.6490 0.0080729 **
AGE         -0.2087772  0.1199920 -1.7399 0.0818721 .
LOTSZ        0.0349699  0.0249192  1.4033 0.1605187
SQFT        -0.0057675  0.2226656 -0.0259 0.9793354
lambda       0.5583108  0.0949300  5.8813 4.071e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The second way is using the function `spregimes` after transforming W into a block-diagonal matrix.

The code below creates two objects of class `listw`, corresponding to the blocks of the spatial weighting matrix. `l0` and `l1` are then transformed in `matrix` and organized to form a block diagonal matrix.

```
R> l0  <- subset(wlis, baltim$CITCOU == 0)
R> l1  <- subset(wlis, baltim$CITCOU == 1)
R> lm0 <- listw2mat(l0)
R> lm1 <- listw2mat(l1)
R> blk <- rbind(cbind(lm0, matrix(0,nrow = nrow(lm0),
ncol = ncol(lm1))),
+                cbind(matrix(0, nrow =
 nrow(lm1), ncol = ncol(lm0)), lm1))
```

The formula was introduced in Sect. 6, and it is written in such a way that all variables are different by regimes. In spregimes the data should be ordered according to the regime variable CITCOU and the argument wy_rg should be TRUE:

```
R> mod_blk_wy <- spregimes(formula = form,
+                          data = baltim[order(baltim$CITCOU),],
+                         rgv = ~ CITCOU, model = "lag",
+                         wy_rg = TRUE, listw = blk, het = TRUE)
R> summary(mod_blk_wy)

     ------------------------------------------------------------
                         Spatial Lag Regimes Model
     ------------------------------------------------------------

Call:
spregimes(formula = form, data = baltim[order(baltim$CITCOU),
   ], model = "lag", listw = blk, wy_rg = TRUE, rgv = ~CITCOU,
   het = TRUE)


Coefficients:
               Estimate Std. Error z-value  Pr(>|z|)
(Intercept)_0  1.2953583 10.5910138  0.1223 0.9026556
NROOM_0        1.4831393  2.0542079  0.7220 0.4702941
NBATH_0        4.5858931  2.1871579  2.0967 0.0360169 *
PATIO_0       11.4988643  5.5027256  2.0897 0.0366477 *
FIREPL_0       6.6285491  4.0309351  1.6444 0.1000895
AC_0          11.9443984  3.0881656  3.8678 0.0001098 ***
GAR_0          0.1726815  2.8685973  0.0602 0.9519986
AGE_0          0.0630254  0.0925082  0.6813 0.4956844
LOTSZ_0        0.1472089  0.0476800  3.0874 0.0020189 **
SQFT_0        -0.1935806  0.3132577 -0.6180 0.5366020
(Intercept)_1  1.4189637  7.4445519  0.1906 0.8488356
NROOM_1        0.5718325  1.4873796  0.3845 0.7006403
NBATH_1       10.3693049  2.9844099  3.4745 0.0005118 ***
PATIO_1        1.1846154  3.1137637  0.3804 0.7036152
FIREPL_1       7.8769605  2.5384508  3.1031 0.0019153 **
AC_1           0.1223055  3.1476928  0.0389 0.9690055
GAR_1          7.9206191  2.9900327  2.6490 0.0080729 **
AGE_1         -0.2087772  0.1199920 -1.7399 0.0818721 .
LOTSZ_1        0.0349699  0.0249192  1.4033 0.1605187
SQFT_1        -0.0057675  0.2226656 -0.0259 0.9793354
W_PRICE_0      0.1949246  0.2163748  0.9009 0.3676599
W_PRICE_1      0.5583108  0.0949300  5.8813 4.071e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Endogenous variables:
W_PRICE_0 W_PRICE_1
```

```
Instruments:
X WX WWX
```

Note that if we do not change the spatial weighting matrix and keep `wlis` (i.e., the queen), we obtain results that are different from column (1) of Table 2.[21]

```
R> mod_blk_wy_2 <- spregimes(formula = form, data = baltim,
+                            rgv = ~ CITCOU, model = "lag",
+                       wy_rg = TRUE, listw = wlis, het = TRUE)
R> summary(mod_blk_wy_2)
     ------------------------------------------------------------
                       Spatial Lag Regimes Model
     ------------------------------------------------------------

Call:
spregimes(formula = form, data = baltim, model =
 "lag", listw = wlis,
   wy_rg = TRUE, rgv = ~CITCOU, het = TRUE)


Coefficients:
                 Estimate Std. Error z-value  Pr(>|z|)
(Intercept)_0 -12.401853  10.701496 -1.1589 0.2465012
NROOM_0         1.903319   2.323058  0.8193 0.4126061
NBATH_0         4.633992   2.477804  1.8702 0.0614559 .
PATIO_0        12.910226   4.789703  2.6954 0.0070302 **
FIREPL_0        3.144275   4.605746  0.6827 0.4948058
AC_0           10.370991   3.247522  3.1935 0.0014055 **
GAR_0           0.560913   2.892655  0.1939 0.8462469
AGE_0           0.087409   0.092747  0.9424 0.3459645
LOTSZ_0         0.109504   0.048979  2.2357 0.0253688 *
SQFT_0         -0.166045   0.396507 -0.4188 0.6753852
(Intercept)_1   3.004709   7.970575  0.3770 0.7061920
NROOM_1         0.926782   1.509713  0.6139 0.5392947
NBATH_1        11.587012   3.131539  3.7001 0.0002155 ***
PATIO_1         3.229213   3.336469  0.9679 0.3331176
FIREPL_1        9.177195   2.719114  3.3751 0.0007380 ***
AC_1            1.049452   3.120094  0.3364 0.7366050
GAR_1           8.072684   3.023883  2.6696 0.0075932 **
AGE_1          -0.182964   0.129269 -1.4154 0.1569590
LOTSZ_1         0.035853   0.025913  1.3836 0.1664766
SQFT_1          0.034014   0.224265  0.1517 0.8794467
W_PRICE_0       0.571567   0.151722  3.7672 0.0001651 ***
W_PRICE_1       0.415025   0.098743  4.2031 2.633e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Endogenous variables:
W_PRICE_0 W_PRICE_1
Instruments:
X WX WWX
```

Column (4) of Table 2 can be matched by estimating the following model:

```
R> mod_blk_nwy <- spregimes(formula = form,
+                           data = baltim[order(baltim$CITCOU),],
```

[21] This is because `spreg.GM_Lag_Regimes` transforms the spatial weighting matrix to a block-diagonal when all variables are different by regimes.

```
+                          rgv = ~ CITCOU, model = "lag",
+                     wy_rg = FALSE, listw = blk, het = TRUE)
R> summary(mod_blk_nwy)

     ------------------------------------------------------------
                     Spatial Lag Regimes Model
     ------------------------------------------------------------

Call:
spregimes(formula = form, data = baltim[order(baltim$CITCOU),
   ], model = "lag", listw = blk, wy_rg = FALSE, rgv = ~CITCOU,
   het = TRUE)


Coefficients:
              Estimate Std. Error z-value  Pr(>|z|)
(Intercept)_0 -8.745818  10.765097 -0.8124 0.4165486
NROOM_0        1.805218   2.318706  0.7785 0.4362476
NBATH_0        4.888737   2.459954  1.9873 0.0468860 *
PATIO_0       12.105956   4.866009  2.4879 0.0128514 *
FIREPL_0       5.401317   4.526198  1.1933 0.2327342
AC_0          11.024414   3.245435  3.3969 0.0006815 ***
GAR_0          0.269524   2.896548  0.0931 0.9258637
AGE_0          0.084914   0.091071  0.9324 0.3511303
LOTSZ_0        0.125088   0.048450  2.5818 0.0098281 **
SQFT_0        -0.240382   0.379826 -0.6329 0.5268166
(Intercept)_1  2.985645   7.476108  0.3994 0.6896293
NROOM_1        0.754364   1.475543  0.5112 0.6091796
NBATH_1       10.830060   3.026499  3.5784 0.0003457 ***
PATIO_1        1.970999   3.186714  0.6185 0.5362426
FIREPL_1       8.373959   2.576870  3.2497 0.0011554 **
AC_1           0.377015   3.153291  0.1196 0.9048298
GAR_1          8.080742   3.029273  2.6676 0.0076406 **
AGE_1         -0.211209   0.123277 -1.7133 0.0866585 .
LOTSZ_1        0.036701   0.024884  1.4749 0.1402478
SQFT_1         0.014546   0.220154  0.0661 0.9473203
W_PRICE        0.480431   0.093026  5.1645 2.411e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Endogenous variables:
W_PRICE
Instruments:
X WX WWX
```

In this model `wy_rg = FALSE`, and the spatial weighting matrix is block diagonal.

The final comparison concerns a model where `wy_rg` is `FALSE`, and the spatial weighting matrix is based on the queen criteria. This model was estimated in Sect. 6, and corresponds to column (2) of Table 2.

```
R> summary(mod_nwy)

    ----------------------------------------------------------
                     Spatial Lag Regimes Model
    ----------------------------------------------------------

Call:
spregimes(formula = form, data = baltim, model =
```

```
 "lag", listw = wlis,
   wy_rg = FALSE, rgv = ~CITCOU, het = TRUE)


Coefficients:
                Estimate Std. Error z-value  Pr(>|z|)
(Intercept)_0 -8.919530  10.948978 -0.8146 0.4152755
NROOM_0         1.924245   2.288908  0.8407 0.4005257
NBATH_0         4.516571   2.422119  1.8647 0.0622208 .
PATIO_0        12.507903   5.227201  2.3928 0.0167181 *
FIREPL_0        3.190648   4.691420  0.6801 0.4964394
AC_0           10.912340   3.053392  3.5738 0.0003518 ***
GAR_0           0.989436   2.880399  0.3435 0.7312175
AGE_0           0.084245   0.091300  0.9227 0.3561486
LOTSZ_0         0.112010   0.048499  2.3095 0.0209156 *
SQFT_0         -0.137437   0.382660 -0.3592 0.7194735
(Intercept)_1  2.329483   7.934102  0.2936 0.7690606
NROOM_1         0.940603   1.555722  0.6046 0.5454389
NBATH_1        11.788691   3.190693  3.6947 0.0002201 ***
PATIO_1         3.576386   3.352183  1.0669 0.2860247
FIREPL_1        9.304493   2.733959  3.4033 0.0006658 ***
AC_1            1.613382   3.203806  0.5036 0.6145545
GAR_1           7.782683   3.039714  2.5603 0.0104572 *
AGE_1          -0.144237   0.128171 -1.1253 0.2604428
LOTSZ_1         0.034213   0.025443  1.3447 0.1787275
SQFT_1          0.025647   0.227200  0.1129 0.9101246
W_PRICE         0.416426   0.098491  4.2281 2.357e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Endogenous variables:
W_PRICE
Instruments:
X WX WWX
```

However, the results obtained from the function `spregimes` are (slightly) different from those in `spreg.GM_Lag_Regimes`.[22]

References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht

Anselin L (2010) Thirty years of spatial econometrics. Pap Reg Sci 89(1):3–25

Anselin L, Rey SJ (2014) Modern spatial econometrics in practice: a guide to GeoDa, GeoDaSpace and PySal. GeoDa Press LLC, Chicago

Anselin L, Amaral P (2021) Endogenous spatial regimes

Aquaro M, Bailey N, Pesaran MH (2020) Estimation and inference for spatial models with heterogeneous coefficients: an application to us house prices. J Appl Econom 36:18–44

Arcaya M, Brewster M, Zigler CM, Subramanian S (2012) Area variations in health: a spatial multilevel modeling approach. Health Place 18(4):824–831. Infectious Insecurities

Arraiz I, Drukker DM, Kelejian HH, Prucha IR (2010) A spatial Cliff-Ord-type model with heteroskedastic innovations: small and large sample results. J Reg Sci 50(2):592–614

[22] We investigated the issue and we found that, while the matrix of regressors is the same, the matrix of instruments is different. In particular, the spatial lag of the dependent variable is calculated using the queen contiguity matrix for both implementations. However, while `spregimes` calculates the lag of the regressors from the same matrix, `spreg.GM_Lag_Regimes` uses a block diagonal matrix.

Baller R, Anselin L, Messner S, Deane G, Hawkins D (2001) Structural covariates of US county homicide rates: incorporating spatial effects. Criminology 39:561–590
Bates D, Maechler M, Jagan M (2022) Matrix: sparse and dense matrix classes and methods. R package version 1.5-1
Bivand RS, Piras G (2015) Comparing implementations of estimation methods for spatial econometrics. J Stat Softw 63(1):1–36
Bivand R, Millo G, Piras G (2021) A review of software for spatial econometrics in R. Mathematics 9(11):1276
Bivand R, Yu D (2022) spgwr: geographically weighted regression. R package version 0.6-35
Chen J, Shin Y, Zheng C (2022) Estimation and inference in heterogeneous spatial panels with a multifactor error structure. J Econom 229(1):55–79
Dahl DB, Scott D, Roosen C, Magnusson A, Swinton J (2019) xtable: export tables to LaTeX or HTML. R package version 1.8-4
Dambon JA, Sigrist F, Furrer R (2021) Joint variable selection of both fixed and random effects for Gaussian process-based spatially varying coefficient models. Int J Geogr Inf Sci 36:2525–2548
Dambon JA, Sigrist F, Furrer R (2021) Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. Spat Stat 41:100470
de Marsily G, Delay F, Gonçalvès J, Renard P, Teles V, Violette S (2005) Dealing with spatial heterogeneity. Hydrogeol J 13:161–183
Deng Y, Qi W, Fu B, Wang K (2020) Geographical transformations of urban sprawl: exploring the spatial heterogeneity across cities in china 1992–2015. Cities 105:102415
Dubir RA (1992) Spatial autocorrelation and neighborhood quality. Reg Sci Urban Econ 22(3):433–452
Ertur C, Le Gallo J, Baumont C (2006) The European regional convergence process, 1980–1995: do spatial regimes and spatial dependence matter? Int Reg Sci Rev 29(1):3–34
Finley AO, Banerjee S, Carlin BP (2007) spBayes: an R package for univariate and multivariate hierarchical point-referenced spatial models. J Stat Softw 19(4):1–24
Finley AO, Banerjee S, Gelfand EA (2015) spBayes for large univariate and multivariate point-referenced spatio-temporal data models. J Stat Softw 63(13):1–28
Fotheringham AS, Brundson C, Charlton M (1998) Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environ Plan A 30:1905–1927
Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted regression. Wiley, Chichester
Fox J, Weisberg S (2019) An R companion to applied regression, 3rd edn. Sage, Thousand Oaks
Geniaux G, Martinetti D (2017) A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. Reg Sci Urban Econ 72:74–85
Genz A, Bretz F (2009) Computation of multivariate normal and t probabilities. Lecture notes in statistics. Springer, Heidelberg
Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2021) mvtnorm: multivariate normal and t distributions. R package version 1.1-3
Georganos S, Grippa T, Niang Gadiaga A, Linard C, Lennert M, Vanhuysse S, Mboga N, Wolff E, Kalogirou S (2021) Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto Int 36(2):121–136
Gollini I, Lu B, Charlton M, Brunsdon C, Harris P (2015) GWmodel: an r package for exploring spatial heterogeneity using geographically weighted models. J Stat Softw
He H, Schäfer B, Beck C (2022) Spatial heterogeneity of air pollution statistics. arXiv preprint arXiv:2203.04296
Huiban J-P, Détang-Dessendre C, Aubert F (2004) Urban versus rural firms: is there a spatial heterogeneity of labour demand? Environ Plan A 36(11):2033–2045
Irwin EG, Bockstael NE (2007) The evolution of urban sprawl: evidence of spatial heterogeneity and increasing land fragmentation. Proc Natl Acad Sci 104(52):20672–20677
Kelejian HH, Piras G (2020) Spillover effects in spatial models: generalizations and extensions. J Reg Sci 60(2):425–442
Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. Int Econ Rev 40:509–533
Kelejian HH, Prucha IR (2007) HAC estimation in a spatial framework. J Econom 140(1):131–154
Kelejian HH, Prucha IR (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. J Econom 157:53–67

Kelejian HH, Prucha IR (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. J Econom 157(1):53–67
LeSage JP, Chih Y-Y (2018) A Bayesian spatial panel model with heterogeneous coefficients. Reg Sci Urban Econ 72:58–73 (**New Advances in Spatial Econometrics: Interactions Matter**)
LeSage J, Pace RK (2009) Introduction to spatial econometrics. Chapman and Hall/CRC, Boca Raton
Lesnoff M, Lancelot R (2022) aods3: analysis of overdispersed data using S3 methods. aods3 package version 0.4-1.2
Longhi S, Nijkamp P (2005) Forecasting regional labour market developments under spatial heterogeneity and spatial autocorrelation. Technical report, Tinbergen Institute Discussion Paper
MATLAB (2011) version 7.13 (R2011b). The MathWorks Inc., Natick, Massachusetts
Melo PC, Graham DJ, Noland RB (2012) The effect of labour market spatial structure on commuting in England and Wales. J Econ Geogr 12(3):717–737
Messner S, Anselin L, Hawkins D, Deane G, Tolnay S, Baller R (2000) An Atlas of the spatial patterning of county-level homicide, 1960–1990. National Consortium on Violence Research (NCOVR), Pittsburgh
Millo G, Piras G (2012) splm: spatial panel data models in R. J Stat Softw 47(1):1–38
Piras G (2010) sphet: spatial models with heteroskedastic innovations in R. J Stat Softw 35(1):1–21
Piras G, Postiglione P (2022) A deeper look at impacts in spatial Durbin model with sphet. Geogr Anal 54(3):664–684
R Development Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0
Ramajo J, Marquez MA, Hewings GJ, Salinas MM (2008) Spatial heterogeneity and interregional spillovers in the European Union: do cohesion policies encourage convergence across regions? Eur Econ Rev 52(3):551–567
Rey SJ, Anselin L (2007) PySAL: a python library of spatial analytical methods. Rev Reg Stud 37:5–27
Rey SJ, Anselin L (2010) Pysal: a python library of spatial analytical methods. In: Fischer MM, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin, pp 175–193
Rey SJ, Janikas MV (2005) Regional convergence, inequality, and space. J Econ Geogr 5(2):155–176
Rey S, Anselin L, Amaral P, Arribas-Bel D, Cortes R, Gaboardi J, Kang W, Knaap E, Li Z, Lumnitz S, Oshan T, Shao H, Wolf L (2022) PySAL ecosystem: philosophy and implementation. Geogr Anal 54(3):467–487
Sarrias M, Piras G (2022) spldv: spatial models for limited dependent variables. R package version 0.1.1
Shu H, Pei T, Song C, Ma T, Du Y, Fan Z, Guo S (2019) Quantifying the spatial heterogeneity of points. Int J Geogr Inf Sci 33(7):1355–1376
Song Y, Wang J, Ge Y, Xu C (2020) An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: Cases with different types of spatial data. GIScience Remote Sens 57(5):593–610
StataCorp (2007) Stata statistical software: release 10. StataCorp LP, College Station
Thomas LJ, Huang P, Yin F, Luo XI, Almquist ZW, Hipp JR, Butts CT (2020) Spatial heterogeneity can lead to substantial local variations in COVID-19 timing and severity. Proc Natl Acad Sci 117(39):24180–24187
van Rossum G (1995) Python Reference manual. CWI report
Vinatier F, Tixier P, Duyck P-F, Lescourret F (2011) Factors and mechanisms explaining spatial heterogeneity: a review of methods for insect populations. Methods Ecol Evol 2(1):11–22
Wheeler D (2022) gwrr: fits geographically weighted regression models with diagnostic tools. R package version 0.2-2
Zhai S, Feng Y, Yan X, Wei Y, Wang R, Li P (2021) Using spatial heterogeneity to strengthen the neighbourhood effects of urban growth simulation models. J Spat Sci 1–19