spsur: An **R** Package for Dealing with Spatial Seemingly Unrelated Regression Models

Román Mínguez 🔾
University of
Castilla-La Mancha

Fernando A. López 🔾
Technical University
of Cartagena

Jesús Mur 🔾
University of
Zaragoza

Abstract

Spatial seemingly unrelated regression (spatial SUR) models are a useful multiequational econometric specification to simultaneously incorporate spatial effects and correlated error terms across equations. The purpose of the **spsur** R package is to supply a complete set of functions to test for spatial structures in the residual of a SUR model; to estimate the most popular specifications by applying different methods and test for linear restrictions on the parameters. The package also facilitates the estimation of so-called spatial impacts, conveniently adapted to a SUR framework. The package includes functions to simulate datasets with the features decided by the user, which may be useful in teaching activities or in more general research projects. The article concludes with a real data application showing the potential that **spsur** has to examine the relation of individual mobility over geographic areas and the incidence of COVID-19 in Spain during the first lockdown.

Keywords: spatial seemingly unrelated regression models, Lagrange multipliers test, maximum likelihood, instrumental variables, panel data, COVID-19, mobility, R package.

1. Introduction

Seemingly unrelated regression models (SUR from now on) are a type of multiequational econometric formulation that gained popularity with the publication of the seminal paper by Zellner (1962). The SUR framework assures increased efficiency by estimating the system jointly rather than by processing each equation separately. These gains explain the relevant position that SUR models occupy in the applied research agenda. However, spatial econometrics is a particular exception: the applied literature in this area is limited, in spite of the extensive methodological research on spatial SUR models (Anselin 1988a,b; Wang and Kockelman 2007; Baltagi and Bresson 2011; Baltagi and Pirotte 2011; Mur, López, and Herrera 2010; Anselin 2016; López, Mínguez, and Mur 2020).

Anselin (1988a) introduced the term *spatial SUR* for the first time in reference to a model made of "*an equation for each time period, which is estimated for a cross section of spatial units*" (page 141). Anselin's approach focuses on the problem of serial dependence in the errors, where each equation stands for a time period. In this work he did not explore estimation techniques, although the likelihood function, the information matrix, and some spatial autocorrelation tests were presented. From this seminal reference in Anselin's book (Anselin 1988a), a relevant number of methodological contributions have been made present, including developing spatial autocorrelation tests and implementing estimation methods.

With respect to the tests of spatial autocorrelation developed in the residual, several papers have focused on this topic. For example, Anselin (1988b) develops a test based on the Lagrange multiplier (LM) principle for spatial error SUR models. Mur *et al.* (2010) design robust and non-robust LM tests including an extensive Monte Carlo exercise for small sample sizes. Baltagi and Bresson (2011) propose joint and conditional LM tests for spatial autocorrelation and random effects for a spatial SUR panel model, while López, Mur, and Angulo (2014) address the problem of selecting the best specification for spatial SUR models using a battery of LM tests.

Several papers have focused on estimation methods. The maximum likelihood (ML) approach under the normality assumption is the most frequently used method, though other alternatives have been proposed. The paper by Wang and Kockelman (2007) makes a significant contribution to this literature, implementing a three-step estimation method combining feasible generalized least squares (FGLS) and ML, which allows for heterogeneity among the individuals by introducing unobserved random effects "à la KKP" (Kapoor, Kelejian, and Prucha 2007). Baltagi and Bresson (2011) extend the results by Wang and Kockelman (2007) by developing a collection of misspecification tests under an ML framework, whereas Baltagi and Pirotte (2011) introduce the generalized method of moments (GMM) in a spatial SUR with spatial autocorrelated errors, also "à la KKP". Anselin (1988a, page 146) describes the three-least-squares (3SLS) estimation method based on instrumental variables (IV) for spatial lag SUR models and Anselin (2016) points out the choice of the instruments. López *et al.* (2020) develop an extensive Monte Carlo exercise to compare ML and 3SLS (based on IV) methods for spatial lag and spatial Durbin SUR models. The method of generalized spatial three-stage least squares (GS3SLS), developed by Kelejian and Prucha (2004) in the more general framework of spatial simultaneous systems, has been used to estimate spatial SUR Models. The GS3SLS method is based on an iterative process that uses IV and GMM (Kelejian and Prucha 1999). Finally, Kakamu, Polasek, and Wago (2012) develop a Bayesian approach for the estimation of spatial SUR models using a Markov chain Monte Carlo (MCMC) method.

Simultaneously, an important number of empirical papers have implemented these methodological developments. Papers by Rey and Montouri (1999), Egger and Pfaffermayr (2004), Le Gallo and Dall'erba (2006), Lundberg (2006), Fingleton (2007), Moscone, Tosetti, and Knapp (2007), Le Gallo and Chasco (2008), Lauridsen, Bech, López, and Maté (2010), Kakamu *et al.* (2012), Cotteleer and Van Kooten (2012), He, Páez, Liu, and Jiang (2015), Izón, Hand, Mccollum, Thacher, and Berrens (2016), López, Martínez-Ortiz, and Cegarra-Navarro (2017), Franklin and Rey (2016), Tupy, Silva, Amaral, and Cavalcante (2021), and Páez, López, Menezes, Cavalcanti, and Pitta (2021) are well-known examples in this line.

From the discussion above, it is clear that there is growing interest in Seemingly Unrelated Regressions in a spatial context, although the lack of specific software has delayed its adoption

in applied research. Empirical applications are still hindered by the lack of readily available user-friendly software. Our purpose in this paper is to introduce the R package called **spsur** to fill this gap. The **spsur** R package offers accessible tools for those researchers who need to incorporate multiequational models into their research when spatial effects and correlation between residuals are present. The **spsur** user has a powerful tool to test for spatial autocorrelation in the residuals of a classic SUR model. This package can estimate spatial SUR models by ML and 3SLS[1], and it uses the Wald statistic to test hypotheses about the stability of parameters, including the parameters of spatial dependence. It can even estimate a restricted model when the parameters are equal in several equations. Direct, indirect, and total impacts can be obtained for all the specifications. Finally, it can estimate uniequational spatial models if it is necessary to compare the estimation of each equation with the result of the multiequational model.

The structure of this paper is the following. Section 2 introduces the notation and presents the main methodology developed on spatial SUR models in the literature. Section 3 presents the structure of the package and some of the main functions included in **spsur**. It focuses on the estimation methods, the misspecification tests, and the computation of spatial impacts (direct, indirect, and total). Section 4 introduces an additional functionality that facilitates simulation experiments that could be helpful in research and/or teaching activities. Section 5 presents an illustration of the **spsur** package using real data. Finally, Section 6 concludes with a brief summary and prospects for the future.

2. Spatial SUR models

The SUR model consists of several equations, possibly with different regressors, where the error terms are correlated. We focus on Anselin's case with a single equation, $G = 1$, several time periods, Tm, and N individuals in each cross section. Usually, $N > Tm$. Originally, the spatial SUR specification assumes that the cross-sectional dimension N is large and that the time dimension Tm is finite. For most regional studies, this assumption of short panels holds true. Following the usual notation, we consider that each equation refers to a time period although the model can be used to estimate different variables in the same cross section.

The baseline model without spatial effects is the following:

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta}_t + \mathbf{u}_t \; ; \qquad \mathsf{E}[\mathbf{u}_t] = 0 \; ; \qquad \mathsf{E}[\mathbf{u}_t \mathbf{u}_s^\top] = \sigma_{ts} \mathbf{I}_N \qquad t, s = 1, \ldots, Tm \qquad (1)$$

where $\mathbf{y_t}$ and $\mathbf{u}_t$ are $N \times 1$ vectors, $\mathbf{X}_t$ is an $N \times p_t$ matrix, with p_t the number of regressors that appear in the t-th equation, and $\boldsymbol{\beta}_t$ is the corresponding $(p_t \times 1)$ vector of coefficients. This model is spatially independent. Therefore, following the terminology in López *et al.* (2014), we call it SUR-SIM.

The serial dependence in the errors of (1) is not parameterized but estimated in the $Tm \times Tm$ covariance matrix, $\boldsymbol{\Sigma} = (\sigma_{ts}; t, s = 1, \ldots, T_m)$.

When we are dealing with spatial data, a frequent problem is spatial dependence, an unobservable feature that can appear in different ways. A general SUR model which includes most

---

[1]The GS3SLS estimation method has been added to the package in the most recent CRAN version 1.0.2.5. We thank an anonymous referee for their suggestion.

of the possible spatial effects of interest in applied research is the following (Elhorst 2014):

$$\mathbf{y}_t = \rho_t \mathbf{W}_t^y \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}_t + \mathbf{W}_t^x \mathbf{X}_t^* \boldsymbol{\theta}_t + \mathbf{u}_t; \ \mathbf{u}_t = \lambda_t \mathbf{W}_t^u \mathbf{u}_t + \boldsymbol{\varepsilon}_t \ ;$$

$$\mathsf{E}[\boldsymbol{\varepsilon}_t] = 0 \ ; \ \ \mathsf{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s^\top] = \sigma_{ts} \mathbf{I}_N \tag{2}$$

where $\mathbf{X}_t^*$ is the matrix of regressors excluding the intercept term; $\mathbf{W}_t^y$, $\mathbf{W}_t^x$, and $\mathbf{W}_t^u$ are $N \times N$ weighting matrices. The model in (2) is referred to as a general nesting model (SUR-GNM, from now on). It is normal to assume that the weighting matrices are the same, $\mathbf{W}_t^x = \mathbf{W}_t^u = \mathbf{W}_t^\varepsilon = \mathbf{W}$. Moreover, the user should supply the exogenous $\mathbf{W}$ matrix based on a priori knowledge (other alternatives appear in Harris, Moffit, and Kravtsova 2011; Kelejian and Piras 2017). This matrix is vital in the specification and must conform to the requisites described in Kelejian and Prucha (2004). The terms on the main diagonal are zero and the row and column sums of $\mathbf{W}$ are uniformly bounded in absolute value, in the same way as $\mathbf{A}_t^{-1} = (\mathbf{I}_N - \rho_t \mathbf{W})^{-1}$ and $\mathbf{B}_t^{-1} = (\mathbf{I}_N - \lambda_t \mathbf{W})^{-1}$, which must exist. The model of (2) can be expressed in matrix form:

$$\mathbf{A}\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_{Tm} \otimes \mathbf{W})\mathbf{X}^* \boldsymbol{\theta} + \mathbf{u}; \ \ \mathbf{B}\mathbf{u} = \boldsymbol{\varepsilon}; \ \ \boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Omega})$$

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \cdots \\ \mathbf{y}_{Tm} \end{bmatrix}_{NTm \times 1}; \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{X}_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{X}_{Tm} \end{bmatrix}_{NTm \times p}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \cdots \\ \boldsymbol{\beta}_{Tm} \end{bmatrix}_{p \times 1}; \ \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \cdots \\ \boldsymbol{\theta}_{Tm} \end{bmatrix}_{(p-Tm) \times 1}; \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \cdots \\ \mathbf{u}_{Tm} \end{bmatrix}_{NTm \times 1}; \ \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \cdots \\ \boldsymbol{\varepsilon}_{Tm} \end{bmatrix}_{NTm \times 1} \tag{3}$$

where $\mathbf{A} = \mathbf{I}_{NTm} - \boldsymbol{\Gamma} \otimes \mathbf{W}$ with $\boldsymbol{\Gamma} = \text{diag}(\rho_1, \ldots, \rho_{Tm})$; $\mathbf{B} = \mathbf{I}_{NTm} - \boldsymbol{\Lambda} \otimes \mathbf{W}$ with $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_{Tm})$. $\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbf{I}_N$, with $\boldsymbol{\Sigma} = [\sigma_{ij}; i, j = 1, \ldots, T_m]$, $\otimes$ denotes the Kronecker product and $p = \sum p_t$. Moreover, $\mathbf{X}^*$ is an $NTm \times (p - Tm)$ that omits the columns of the intercepts (these terms cannot be spatially lagged).

Several constrained specifications emerge from (3), such as:

1. The SUR-SIM spatially independent model, $\rho_t = \lambda_t = 0; \boldsymbol{\theta}_t = \mathbf{0} \ \ (\forall t)$.

2. The SUR-SLX spatial lags in the Xs model, where $\rho_t = \lambda_t = 0 \ \ (\forall t)$.

3. The SUR-SLM spatial lag model, where $\lambda_t = 0; \boldsymbol{\theta}_t = \mathbf{0} \ \ (\forall t)$.

4. The SUR-SEM spatial error model, where $\rho_t = 0; \boldsymbol{\theta}_t = \mathbf{0}, \ \ (\forall t)$.

5. The SUR-SDM spatial Durbin model, where $\lambda_t = 0 \ \ (\forall t)$.

6. The SUR-SDEM spatial Durbin error model, where $\rho_t = 0 \ \ (\forall t)$.

7. The SUR-SARAR spatial lag model with autoregressive errors, where $\boldsymbol{\theta}_t = \mathbf{0} \ \ (\forall t)$.

The spatial SUR model specified in (3) can be easily extended to multi-dimensional panels with $G > 1$ (Wang and Kockelman 2007; López *et al.* 2014), which is omitted here for simplicity. This multi-dimensional panel model with $N > 1$, $Tm > 1$, and $G > 1$ has been used in López *et al.* (2014) to compare model selection strategies in the context of spatial SUR models, in López *et al.* (2020) to compare ML and 3SLS (based on IV) methods to estimate some types of spatial SUR models, and it has been applied by López *et al.* (2017) to analyze spatial spillovers in the public expenditures of local governments at a municipal level.

2.1. Testing for spatial effects

The Lagrange multipliers presented below test whether the simple nonspatial model of (1) is misspecified due to omitted, unobserved spatial factors. All of them are obtained in an ML framework. The general expression of the multipliers, including a detailed description of the evaluation process, can be found in Mur *et al.* (2010). First, we present a global test for the presence of spatial lags both in the SUR equation and in the errors:

$$H_0 : \rho_t = \lambda_t = 0 \ \ (\forall t) \quad \text{vs} \quad H_A : \exists t \text{ with } \rho_t \neq 0 \ \text{ and } \lambda_t \neq 0 \tag{4}$$

The model of the null is the SUR-SIM of (1), whereas the alternative is the SUR-SARAR. The corresponding Lagrange multiplier is called $\text{LM}_{\text{SARAR}}^{\text{SUR}}$. The next multiplier ($\text{LM}_{\text{SLM}}^{\text{SUR}}$) maintains the model of the null, which is the SUR-SIM, but that of the alternative is the SUR-SLM, so:

$$H_0 : \rho_t = 0 \ \ (\forall t) \quad \text{vs} \quad H_A : \exists t \text{ with } \rho_t \neq 0 \tag{5}$$

Implicitly, it is assumed that the λ parameters are zero. Finally, the $\text{LM}_{\text{SEM}}^{\text{SUR}}$ places a SUR-SEM model in the alternative against, once more, the SUR-SIM in the null (assuming that the ρ parameters are zero):

$$H_0 : \lambda_t = 0 \ \ (\forall t) \quad \text{vs} \quad H_A : \exists t \text{ with } \lambda_t \neq 0$$

Bera and Yoon (1993) show that the $\text{LM}_{\text{SLM}}^{\text{SUR}}$ and $\text{LM}_{\text{SEM}}^{\text{SUR}}$ multipliers are oversized in the case of misspecifying the alternative hypothesis (i.e., we want to test the SUR-SIM model in the null using the $\text{LM}_{\text{SLM}}^{\text{SUR}}$ of (5) when, in fact, the data have been generated by a SUR-SEM). This lack of robustness makes the identification of the correct spatial specification more complicated. The solution, as suggested in Anselin, Bera, Florax, and Yoon (1996), is to *robustify* the *raw* multipliers using the so-called robust Lagrange multipliers, denoted as $\text{LM}_{\text{SLM}}^{*\text{SUR}}$ and $\text{LM}_{\text{SEM}}^{*\text{SUR}}$, respectively. The size of the robust multipliers is more balanced at the cost of a slight loss of power.

The LM tests can be used as a guide to select the correct specification in the case of spatial SUR models (López *et al.* 2014) similar to those that have been used in a cross section framework (Mur and Angulo 2009), but problems can occur under certain conditions, as explained by Piras and Prucha (2014). Alternatively, the user could decide on the most suitable specification based on deep knowledge about the problem and not on the pre-test results (for example, guided by economic theory).

2.2. Estimation of spatial SUR models

The estimations of spatial SUR models are a direct extension of the methodology used to estimate single, spatial cross-section models. In the case of spatially lagged variables, the

endogeneity of the spatial lag must be dealt with and, in the case of spatially correlated errors, the non-spherical nature of the covariance matrix must be accounted for. Several estimation approaches have been suggested in the literature. The first is based on the ML principle under the hypothesis of normality. The main problem with this method is that it is highly demanding in terms of computational effort when the size of the cross section, N, increases due to the Jacobian determinant in the concentrated likelihood function. The second approach is based on the methods of 3SLS that use IV and on GMM. These methodologies are alternatives to the often unrealistic assumption of normality, and they help to avoid computational problems with big datasets. A review of these methodologies in the framework of spatial panel data, including spatial SUR models, is found in Anselin, Le Gallo, and Jayet (2008), and we present a condensed version for the specific case of spatial SUR models in this section.

The maximum likelihood estimation

Inference based on the ML method has been the most commonly used method to obtain the estimation of the coefficients of a spatial SUR model (e.g., Wang and Kockelman 2007; Baltagi and Bresson 2011; Lauridsen *et al.* 2010; Le Gallo and Dall'erba 2006; López *et al.* 2017). In this case, assuming that the errors are normally distributed, the log-likelihood function of (3) can be written as:

$$\ell(\mathbf{y}; \boldsymbol{\eta}) = -\frac{NTm}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Sigma}| + Tm\sum_{t=1}^{Tm}\ln|\mathbf{B}_t| + Tm\sum_{t=1}^{Tm}\ln|\mathbf{A}_t|$$
$$- \frac{\left(\mathbf{Ay} - \bar{\mathbf{X}}[\boldsymbol{\beta};\ \boldsymbol{\theta}]^\top\right)^\top \mathbf{B}^\top \boldsymbol{\Omega}^{-1} \mathbf{B}\left(\mathbf{Ay} - \bar{\mathbf{X}}[\boldsymbol{\beta};\ \boldsymbol{\theta}]^\top\right)}{2} \quad (6)$$

where $\bar{\mathbf{X}} = [\mathbf{X}, (\mathbf{I}_{Tm} \otimes \mathbf{W})\mathbf{X}^*]$ is an $NTm \times (2p - Tm)$ matrix.

The vector of parameters is $\boldsymbol{\eta}^\top = \left[\boldsymbol{\beta}^\top; \boldsymbol{\theta}^\top; \rho_1; \ldots; \rho_{Tm}; \lambda_1; \ldots; \lambda_{Tm}; \sigma_{ij}\right]$. In total, there are $P = (2p - Tm) + 2Tm + Tm(Tm + 1)/2$ unknown coefficients. Wang and Kockelman (2007) and López *et al.* (2014) use numerical optimization techniques to solve the maximum likelihood estimation. Note that the log of the Jacobian terms, $\ln|\mathbf{B}_t|$ and $\ln|\mathbf{A}_t|$, constrains the spatial parameters, ρ_t and λ_t, to lie inside the so-called *stability interval* (if the weighting matrix has been row-standardized, a rough approximation is $\left(\frac{1}{\delta_{\text{MIN}}^-}; +1\right)$ with δ_{MIN}^- being the greatest negative eigenvalue of $\mathbf{W}$). In practice, estimation consists of applying a non-linear optimization to the concentrated log-likelihood function. Under the assumption that the spatial SUR model is correctly specified, the ML estimators will be consistent, efficient, and asymptotically normally distributed (Davidson and MacKinnon 1993).

A main obstacle in the practical implementation of the ML estimation is the need to compute a Jacobian determinant for an N-dimensional matrix. The classic solution to this problem is to decompose the Jacobian in terms of the eigenvalues of the spatial weights matrix. Other methods are based on Cholesky or lower-upper (LU) decomposition, which exploits the sparsity of the spatial weights matrix $\mathbf{W}$ (see the references included in Bivand, Hauke, and Kossowski 2013a and Piras 2010 about this problem).

The 3SLS estimation method based on instrumental variables

Due to the highly intensive processing requirements of the ML method when N increases and to the often unrealistic assumption of normality, other methods have been proposed to

estimate spatial SUR models. In the case of SUR-SLM and SUR-SDM, the 3SLS method based on instrumental variables (IV) can be used to estimate these models. In these models, the presence of the spatially lagged dependent variables introduces a type of endogeneity that calls for an instrumental variable approach. Anselin (1988a, page 147) and Anselin *et al.* (2008) proposed this methodology, although it was not implemented. Anselin (2016) later proposed this same estimation method and described the adequate instruments. López *et al.* (2020), Tupy *et al.* (2021), and Páez *et al.* (2021) applied this methodology to estimate SUR-SLM models. The list of proposed instruments for SUR-SLM and SUR-SDM, along with a full description of the method, is available in López *et al.* (2020) where an extensive Monte Carlo evaluates this method and compares the results with the ML estimation.

Note that this estimation procedure is almost linear, which simplifies computational effort. On the negative side, we should keep in mind that the estimation of the ρ parameters is not restricted by the Jacobian term, so we can receive estimates of ρ well outside the *stability interval* (although the risk seems acceptable, see López *et al.* 2020).

2.3. Testing for linear restrictions in spatial SUR models

In this section we present the Wald test approach to test linear restrictions on the parameters of an estimated spatial SUR model. Specifically, we present a Wald test to check linear restrictions on the β parameters or to test linear restrictions on the spatial parameters (λ or ρ). Several papers have used the Wald test in spatial SUR models. For example, this test is applied in Cotteleer and Van Kooten (2012) and in He *et al.* (2015) to evaluate significant differences in the β coefficients. Páez *et al.* (2021) apply the Wald test to evaluate the null hypothesis of identical levels of spatial dependence in the ρ_t parameters.

Anselin (2016) describes the Wald test for β coefficients in a spatial SUR framework. Assuming that the model is correctly specified (emphasizing the importance of the misspecification tests discussed in Section 2.1), the ML and 3SLS estimates are asymptotically normally distributed. So, for the ML case, we can write (Lee 2004; Elhorst 2014):

$$\hat{\boldsymbol{\eta}} \underset{as}{\sim} N\left(\boldsymbol{\eta}; \mathbf{V}(\boldsymbol{\eta})\right) \tag{7}$$

where $\boldsymbol{\eta}$ is the vector of the parameters, as defined in (6), and $\mathbf{V}(\boldsymbol{\eta})$ is the corresponding covariance matrix, as described in López *et al.* (2014). A similar result can be stated for the 3SLS case, for which we need the orthogonality conditions defined in Section 2.2.2 and the covariance matrix, whose details can be found in Wooldridge (2010, Chapter 7).

Using the asymptotic distribution of (7), we can test for any kind of linear restriction on the parameter vectors of the model, such as exclusion restrictions or cross-equation linear restrictions. The procedure is carried out in the usual way. First, we need to write down the $(r \times P)$ matrix $\mathbf{R}$, with $r = \text{rank}(\mathbf{R})$, which captures the linear restrictions:

$$H_0: \quad \mathbf{R}\boldsymbol{\eta} = \mathbf{b} \quad vs \quad H_A: \quad \mathbf{R}\boldsymbol{\eta} \neq \mathbf{b} \tag{8}$$

The Wald statistic for testing (8) is

$$W = \left(\mathbf{R}\hat{\boldsymbol{\eta}} - \mathbf{b}\right)^{\top} \left(\mathbf{R}\hat{\mathbf{V}}(\boldsymbol{\eta})\mathbf{R}^{\top}\right)^{-1} \left(\mathbf{R}\hat{\boldsymbol{\eta}} - \mathbf{b}\right) \underset{as}{\sim} \chi_r^2$$

Note that the restrictions on (8) must involve only βs or spatial parameters. With slight variations, we obtain a similar result for the 3SLS method. Alternatively, the LR approach could be used but it is a bit more costly in terms of computing time than its Wald counterpart.

2.4. Direct, indirect, and total impacts

A capital difference between time series and spatial models lies in the interpretation of the $\boldsymbol{\beta}$ parameters. In a time series context, the $\boldsymbol{\beta}$ parameters have an unequivocal meaning: they measure the expected increase in the explained variable as a consequence of a unitary change in a given regressor. This impact is constant across time, $\frac{\partial y_t}{\partial x_t} = \beta_k$, and its significance can be checked by a simple t ratio.

In a spatial setting, there are feedback effects. This means that a change in a regressor in a given location will trigger a chain of reactions that will spread all over the space, according to matrix $\mathbf{W}$. Consider the case of a SUR-SLM model with only one equation, $G = 1$, whose reduced form can be written as:

$$\mathbf{y} = \mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{A}^{-1}\mathbf{u} = \mathbf{I}_{Tm} \otimes \left[\mathbf{I}_N + \boldsymbol{\Lambda} \otimes \mathbf{W} + \boldsymbol{\Lambda}^2 \otimes \mathbf{W}^2 + \ldots\right] \mathbf{X}\boldsymbol{\beta} + \mathbf{A}^{-1}\mathbf{u} \qquad (9)$$

$\mathbf{A}$ is a block-diagonal matrix whose non-zero matrices are $\mathbf{A}_t = \mathbf{I}_N - \rho_t \mathbf{W}; t = 1, 2, \ldots Tm$. Assuming that all the ρ parameters pertain to the so-called *stability interval*, $\mathbf{A}_t^{-1}$ admits a power series expansion in terms of the spatial coefficients and the weighting matrix, which leads us to expression (9). Clearly, if we change any regressor in any part of the space, its effects will impact the explained variable wherever it is located. This is the key starting point for the spatial multiplier analysis. Unfortunately, each type of spatial model produces its own multiplier effects, (LeSage and Pace 2009), so it is difficult to offer general results apart from basic notions and definitions.

Using the classic notation in LeSage and Pace (2009), we can write the following for the t-th period of a SUR-SLM model:

$$\mathbf{y}_t = \begin{bmatrix} y_{t1} \\ y_{t2} \\ \ldots \\ y_{tN} \end{bmatrix} = \sum_{h=1}^{p} \begin{bmatrix} S_t^h(\mathbf{W})_{11} & S_t^h(\mathbf{W})_{12} & \cdots & S_t^h(\mathbf{W})_{1N} \\ S_t^h(\mathbf{W})_{21} & S_t^h(\mathbf{W})_{22} & \cdots & S_t^h(\mathbf{W})_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ S_t^h(\mathbf{W})_{N1} & S_t^h(\mathbf{W})_{N2} & \cdots & S_t^h(\mathbf{W})_{NN} \end{bmatrix} \begin{bmatrix} x_{th1} \\ x_{th2} \\ \ldots \\ x_{thN} \end{bmatrix} + \mathbf{A}^{-1} \begin{bmatrix} u_{t1} \\ u_{t2} \\ \ldots \\ u_{tN} \end{bmatrix}$$

$$(10)$$

We assume that each equation has the same p regressors. $S_t^h(\mathbf{W})_{ij}$ is the (i, j) element of the $(N \times N)$ matrix $S_t^h(\mathbf{W})$ obtained as $S_t^h(\mathbf{W}) = \beta_h \mathbf{A}_t^{-1}$, β_h is the h-th β parameter, and x_{thn} refers to regressor h in period t in region n, whereas y_{tn} is the observed value of the explained variable in period t and region n. Note that unless there are changes in the β parameters or in the weighting matrices $\mathbf{W}$, the multiplier matrices $S_t^h(\mathbf{W})$ remain constant across time.

Expression (10) is crucial because it helps us to map the chain of feedback effects inherent in the specification. For example, the impact of a marginal change in regressor h $(h = 1, \ldots, p)$ in region n on the explained variable located in region s is the (n, s) element of this matrix, $\frac{\partial y_{ts}}{\partial x_{thn}} = S_t^h(\mathbf{W})_{ns}$.

Applied literature distinguishes three types of impacts: The so-called *direct impact*, where all the changes occur in the same region and coincide with the mean of the terms in the main diagonal of the multiplier matrix of (10), *indirect impact* which measures spill-over effects, evaluated as the mean of the off-diagonal terms of $S_t^h(\mathbf{W})$, and *total impact*, as the sum of

the two effects. In an obvious notation:

$$
\begin{aligned}
\mathbf{M}_{T_g^h} &= \frac{1}{N} \tau_N^\top \mathbf{S}_{.g}^h \tau_N \\
\mathbf{M}_{D_g^h} &= \frac{1}{N} \mathrm{tr} \mathbf{S}_{.g}^h \\
\mathbf{M}_{I_g^h} &= \mathbf{M}_{T_g^h} - \mathbf{M}_{D_g^h} \ ; \ (h = 1, \ldots, p_g; \ g = 1, \ldots, G)
\end{aligned}
\tag{11}
$$

with τ_N as an $N \times 1$ unitary vector and $\mathrm{tr}()$ as the trace operator. The subindex g denotes that these effects correspond to the g-th equation of the SUR (if $G = 1$, the subindex g should be dropped).

Finally, for the case of $G > 1$, we can aggregate the multipliers in (11) to obtain overall measures for each variable in the SUR:

$$
\mathbf{M}_{D^h} = \sum_{g=1}^{G} \mathbf{M}_{D_g^h} \ ; \ \ \mathbf{M}_{I^h} = \sum_{g=1}^{G} \mathbf{M}_{I_g^h} \ ; \ \ \mathbf{M}_{\mathrm{Total}^h} = \sum_{g=1}^{G} \mathbf{M}_{\mathrm{Total}_g^h} \ ; \ \ (h = 1, \ldots p)
\tag{12}
$$

The multipliers of (12) should be accompanied by measures of statistical significance to be useful. To simplify, let us assume that the same list of regressors appears in the G equations of the SUR. LeSage and Pace (2009) suggest two possibilities. The first is the Bayesian MCMC approach, which draws samples of parameters from their posterior distribution and computes random sequences of multipliers to calibrate the estimates. To our knowledge, there are no results available on applying MCMC in a SUR framework. The second alternative, implemented in **spsur**, is simulation using the maximum likelihood or 3SLS estimates, which involves the following procedure:

- Estimate the spatial SUR model either by ML (preferably) or 3SLS.

- Draw B samples of the random vectors of the parameters for our spatial SUR, $\boldsymbol{\eta}$, using the estimated multivariate normal distribution, $N\left(\hat{\boldsymbol{\eta}}; \hat{\mathbf{V}}(\hat{\boldsymbol{\eta}})\right)$, where $\hat{\boldsymbol{\eta}}$ and $\hat{\mathbf{V}}(\hat{\boldsymbol{\eta}})$ are their respective estimates. $\boldsymbol{\eta}_b$ denotes the b-th draw.

- Re-evaluate the effects using the simulated vectors $\boldsymbol{\eta}_b$ and maintaining the data of the regressors. In this way, a sequence of values: $M_{\mathrm{Total}_g^{h(b)}}$, $M_{\mathrm{Direct}_g^{h(b)}}$ and $M_{\mathrm{Indirect}_g^{h(b)}}$ for $b = 1, \ldots B$ is obtained.

- Finally, calculate the corresponding dispersion measure as the standard deviation of the series of estimated multipliers:

$$
\sigma_{M_{\mathrm{Effect}_g^h}} = \sqrt{\frac{1}{B} \sum_{b=1}^{B} \left(M_{\mathrm{Effect}_g^{h(b)}} - \overline{M_{\mathrm{Effect}_g^b}} \right)^2}
$$

with Effect = Total, Direct, Indirect.

2.5. Software review

In the field of quantitative analysis of spatial data, we can cite several R specialized packages. The most popular are **sf** (Pebesma 2018), **spdep** (Bivand, Pebesma, and Gómez-Rubio 2013b), **spatialreg** (Bivand and Piras 2015), **sphet** (Piras 2010), and **splm** (Millo and Piras 2012).

spdep and **spatialreg** are two powerful packages with high functionalities for the spatial econometric analysis of cross-sectional data, whereas **spml** is focused on the area of spatial panels. On the other hand, the R package **systemfit** deals with SUR models but does not incorporate spatial effects. To our knowledge, only the embryonic package, **spse** (Piras 2018), downloadable from the R-Forge repository in `https://rdrr.io/rforge/spse`, offers some functionalities to estimate spatial SUR models (see Bivand, Millo, and Piras 2021 for a review of software for spatial econometrics in R). A comparative vignette of the common functionalities of **spsur** and **spse** is available in the GitHub repository at `https://github.com/rominsal/spsur` for the development version of the **spsur** package. In summary, none of these packages are well-equipped to deal with the overall inference problems of spatial SUR models.

The most similar antecedent to **spsur** is the beta version of **SpaceStat** (Anselin 1992) called **SpaceStat** (version May91), with limited functionalities. Moreover, it requires Windows 98 (see the YouTube video: `https://www.youtube.com/watch?v=6pM0tDWqt3o&t=53s`). Of course, nowadays it would be very difficult to do applied research using **SpaceStat** (version May91). There is a second alternative in Python, in the **PySAL** library (Rey and Anselin 2009), which is the module `spreg.sur`. It provides SUR estimations of models with no spatial impacts and diagnostic measures, among which we find the *raw* Lagrange multipliers. Moreover, several functions provide the estimation of spatial SUR models with spatial errors or spatial lags of the explained variable, respectively. The SUR-SEM model is estimated by ML, while the SUR-SLM is estimated using spatial instrumental variables (3SLS based on IV). Comparison of the results of an estimated spatial SUR model with **PySAL** and **spsur** and **spse** is available in a vignette of **spsur**. Finally, there is some code in MATLAB connected to the work by López *et al.* (2014), but the source is not free software.

3. Introducing the spsur package

spsur (López, Mínguez, and Mur 2022) is an open-source software for the R computing platform (R Core Team 2022). The package can be freely downloaded from the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=spsur`. Users will find complete assistance documentation and some vignettes to guide them through the functionalities of the package, as well as examples of its use.

spsur is made of 10 different functions that allow for a thorough analysis of a spatial SUR specification. Figure 1 presents a graphic description of the package. **spsur** includes one function to evaluate the Lagrange multipliers for the presence of spatial correlation in a SUR-SIM model, `lmtestspsur()`. For the general case of $G > 1$, `spsurml()` and `spsur3sls()` allow the estimation of the model by ML and 3SLS (based on IV). The function `spsurtime()` is directed specifically toward spatial SUR models, where $G = 1$ and $Tm > 1$. Four functions are designed to improve specification: `anova()`, `wald_betas()`, `wald_deltas()`, and `lr_betas()`; while `impactspsur()` computes the direct, indirect, and total spatial effects. Finally, `dgp_spsur()` allows the user to solve Monte Carlo experiments using spatial SUR models.

The functions to estimate spatial SUR models `spsurml()` and `spsur3sls()` return an object of the class 'spsur'. This class has been specifically defined in the package and several methods have been implemented to work with this object. The method `summary()` presents a full output of the estimated model, `print()` presents a short output of the results of the
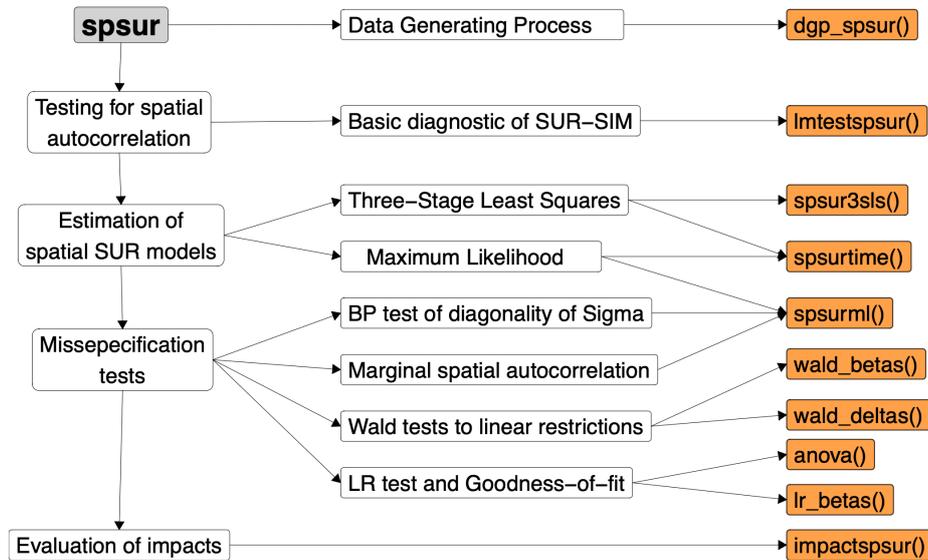
Figure 1: Main functionalities of **spsur** package.

estimation, and `plot()` generates a plot with the coefficients. To conclude, it is important to point out that **spsur** includes several vignettes to guide the user. One vignette shows the main functionalities of the package and another presents some results from the use of **spsur** for single equation and baseline models. Two additional vignettes, available in the `GitHub` version, present a sort of Monte Carlo exercise and a comparison with **spse** and **PySAL**.

The package is installed from the CRAN repository and the dataset is loaded as usual,

```
R> if (!require("spsur")) install.packages("spsur")
R> library("spsur")
```

3.1. Datasets and baseline model

Three datasets have been included in the package; namely, `spc`, `NCOVR` and `spain.covid`. The first two are well known in the spatial econometric literature and are used in this paper to show the performance **spsur**. The third is included in Section 5 to illustrate an empirical case where the spatial SUR approach is useful in econometric methodology. Each dataset consists of an 'sf' object (Pebesma 2018), allowing the construction of a list of neighbors and plot maps.

The `spc` dataset: The first dataset, `spc`, constitutes a classic example of a Phillips curve, taken from Anselin (1988a, pages 203–211). This dataset includes an 'sf' object of 25 counties in Southwest Ohio with information about different variables for two time periods, 1981 and 1983. The explained variable measures the changes in wage rates (WAGE), while the regressors are unemployment rate (UN), net-migration rate (NMR), and a dummy variable (SMSA) with a value of 1 for metropolitan counties.

The SUR-SIM model estimated by Anselin (1988a) is:

$$\begin{aligned}
\text{WAGE}_{83} &= \beta_{10} + \beta_{11}\,\text{UN}_{83} + \beta_{12}\,\text{NMR}_{83} + \beta_{13}\,\text{SMSA} + \varepsilon_{83} \\
\text{WAGE}_{81} &= \beta_{20} + \beta_{21}\,\text{UN}_{80} + \beta_{22}\,\text{NMR}_{80} + \beta_{23}\,\text{SMSA} + \varepsilon_{81}
\end{aligned} \tag{13}$$

We use the **Formula** package (Zeileis and Croissant 2010) to specify the multiequational SUR model of (13):

```
R> spcformula <- WAGE83 | WAGE81 ~ UN83 + NMR83 + SMSA | UN80 + NMR80 + SMSA
```

Note that we first select the two dependent variables appearing in each equation, separated by a vertical bar (|). Next, after the prime, we specify the regressors for each equation, once more using a vertical bar to separate the two groups.

The NCOVR dataset: The second dataset (NCOVR) contains data downloaded from the GeoDa Data and Lab collection about homicide rates in 3,085 continental US counties for four years (1960, 1970, 1980, and 1990). The dataset includes a large number of socio-economic characteristics for these counties. This dataset comes from the paper by Baller, Anselin, Messner, Deane, and Hawkins (2001)[2].

```
R> data("NCOVR", package = "spsur")
```

Following Baller *et al.* (2001), we choose a **W** matrix based on the *k*-nearest-neighbors, with $k = 10$, using the geographic centroids of US counties.

```
R> library("spdep")
R> library("sf")
R> co <- sf::st_coordinates(sf::st_centroid(NCOVR.sf))
R> ncovrlw <- nb2listw(knn2nb(knearneigh(co, k = 10, longlat = TRUE)),
+    style = "W")
```

We consider the next SUR-SIM model with three equations and different numbers of regressors in each one only to show the functionalities of **spsur**:

$$\begin{aligned}
\text{HR}_{80} &= \beta_{10} + \beta_{11}\,\text{PS}_{80} + \beta_{12}\,\text{UE}_{80} + \varepsilon_{\text{HR}} \\
\text{DV}_{80} &= \beta_{20} + \beta_{21}\,\text{PS}_{80} + \beta_{22}\,\text{UE}_{80} + \beta_{23}\,\text{SOUTH} + \varepsilon_{\text{DV}} \\
\text{FP}_{79} &= \beta_{30} + \beta_{31}\,\text{PS}_{80} + \varepsilon_{\text{FP}}
\end{aligned} \tag{14}$$

HR_{80} is the homicide rate per 100,000 inhabitants in 1980, PS_{80} measures the population structure in 1980, UE_{80} is the unemployment rate, DV_{80} is the divorce rate, FP_{79} is the percentage of families below the poverty line in 1980, and SOUTH is a dummy variable for Southern counties. Note that the model of (14) has only one cross section.

To input this model into R, we write,

```
R> ncovrformula <- HR80 | DV80 | FP79 ~
+    PS80 + UE80 | PS80 + UE80 + SOUTH | PS80

old-style crs object detected; please recreate object with a recent sf::st_crs()
old-style crs object detected; please recreate object with a recent sf::st_crs()
old-style crs object detected; please recreate object with a recent sf::st_crs()
```
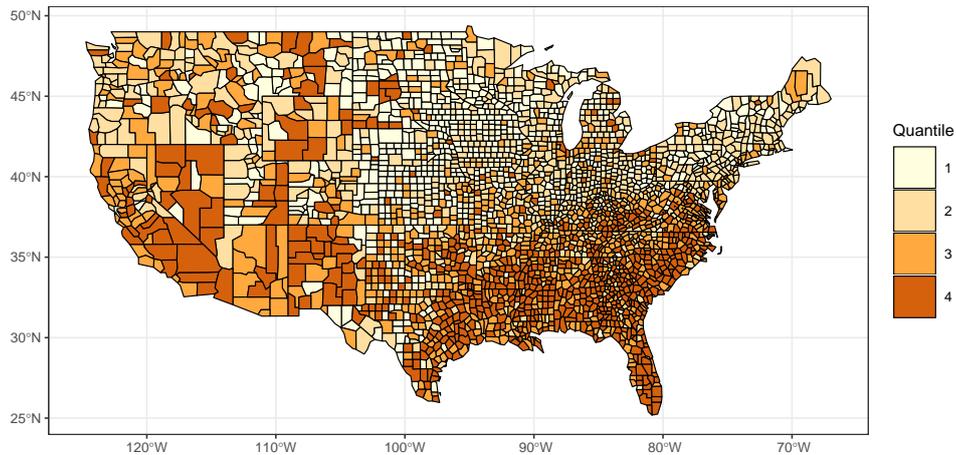
Figure 2: Spatial distribution of homicide rates in US counties in 1980 (HR80).

Figure 2 shows the spatial distribution of homicide rates, in 1980, where a clear spatial structure emerges. Therefore, we expect to find spatial autocorrelation at least in the first equation.

3.2. Misspecification tests for omitted spatial effects

The generic function `lmtestspur()` computes five Lagrange multipliers for omitted spatial dependence in a SUR-SIM model. The main advantage of the LM tests is that they only need information of the null hypothesis, and it is therefore easy to obtain. The full syntax of the first one is:

```
lmtestspsur(formula, data, listw, na.action, time = NULL, Tm = 1,
   zero.policy = NULL, R = NULL, b = NULL, ...)
```

The `lmtestspur()` function includes a 'formula' object, a data frame in `data` argument, a list of neighbors in a 'listw' object (it could also be a **W** neighbor matrix), `zero.policy`, and `na.action` for the treatment of missing values (see Bivand and Wong 2018, for details). Check the help area of the function for the rest of the arguments.

Continuing with the `NCOVR` case above, we should write:

```
R> LMs.ncovr <- lmtestspsur(formula = ncovrformula, data = NCOVR.sf,
+    listw = ncovrlw)
R> pr.LMs.ncovr <- sapply(LMs.ncovr, print)


    LM-SUR-SLM

data:  NCOVR.sf
LM-stat = 8800.7, df = 3, p-value < 2.2e-16
```

---

[2]See https://geodacenter.github.io/data-and-lab/ncovr/ for more details about the database.

```
    LM-SUR-SEM

data:  NCOVR.sf
LM-stat = 9912.7, df = 3, p-value < 2.2e-16


    LM*-SUR-SLM

data:  NCOVR.sf
LM-stat = 19.774, df = 3, p-value = 0.0001891


    LM*-SUR-SEM

data:  NCOVR.sf
LM-stat = 1131.8, df = 3, p-value < 2.2e-16


    LM-SUR-SARAR

data:  NCOVR.sf
LM-stat = 9932.5, df = 6, p-value < 2.2e-16
```

The output of the `lmtestspsur()` function includes a list of 'htest' class objects for each of the LM tests. The situation depicted by this battery of multipliers shows that the SUR-SIM model of (14) is clearly misspecified, but the set of LMs do not reveal the type of misspecifications that afflict the model.

3.3. Maximum Likelihood estimation of spatial SUR models

A spatial SUR model must be estimated in case the LM tests reject the SUR- SIM alternative. The function `spsurml()` obtains the ML estimation of the different spatial SUR models listed in Section 2. The general syntax for this function is:

```
spsurml(formula = NULL, data = NULL, na.action, listw = NULL, type = "sim",
    Durbin = NULL, method = "eigen", zero.policy = NULL, interval = NULL,
    trs = NULL, R = NULL, b = NULL, X = NULL, Y = NULL, G = NULL, N = NULL,
    Tm = NULL, p = NULL, control = list() )
```

Example with `spc`

The symptoms of spatial misspecification in the Phillips curve are rather weak. The user can obtain all the LM tests with the function `lmtestspsur()` shown in Section 3.2, which we omit to save space. However, as an example, suppose we conclude that the SUR-SIM model is misspecified because we have omitted a spatial lag of the explained variables on the right side of the equation (that is, we interpret that the correct specification is a SUR-SLM model, which is plausible if we fix the significance level at 10%).

To estimate a SUR-SLM model for the `spc` dataset, the syntax is:

```
R> spcsur.slm <- spsurml(formula = spcformula, data = spc, type = "slm",
+    method = "eigen", listw = Wspc)


neighbourhood matrix eigenvalues
Computing eigenvalues ...

Initial point:    log_lik:  113.198   rhos:   -0.472 -0.446
Iteration:  1    log_lik:  114.088   rhos:   -0.506 -0.482
Iteration:  2    log_lik:  114.098   rhos:   -0.506 -0.482
Iteration:  3    log_lik:  114.099   rhos:   -0.505 -0.482
Time to fit the model:  1.548   seconds
Time to compute covariances:  0.908   seconds
```

The `listw` argument allows the specification of neighbors either as a list or as a matrix. The `method` argument (the same as the argument of the estimation functions in the **spatialreg** package) permits the choice of method to compute the determinant of the spatial Jacobian in the log-likelihood. For relatively small samples (hundreds of observations), the simplest method is `"eigen"` (which is also the default). Nevertheless, the computation of the eigenvalues becomes very expensive for medium to large samples. The function `spsurml()` offers other options to obtain the spatial determinant without computing the eigenvalues; for example, the Cholesky or LU factorization (`"Matrix"` or `"LU"`), or the Chebyshev approximation (`"Chebyshev"`). See Bivand *et al.* (2013a) and Chapter 4 in LeSage and Pace (2009) for details. The `zero.policy`, `interval`, and `trs` arguments are also similar to those available in the **spatialreg** package (see `?spatialreg::lagsarlm` for the details of each one).

The `type` argument selects the required spatial SUR specification among the available alternatives: `"sim"`, `"slx"`, `"slm"`, `"sem"`, `"sdm"`, `"sdem"`, `"sarar"`, or `"gnm"` (see Section 2 for details of each case). The `Durbin` argument allows the specification of `"sdm"` models, (in addition to `"slx"`, `"sdem"` and `"gnm"`), with only part of the regressors spatially lagged.

Finally, the options included in the `control` argument are almost identical to those available in the `lagsarlm()` function of **spatialreg** (see `?spsurml` for details of all of them). Some of these options include the possibility of checking for intermediate results during the optimization process using `trace = TRUE`, setting the threshold for the convergence of the log-likelihood in `tol`, the maximum number of iterations in `maxit`, and the possibility of computing the numerical covariances using `fdHess = TRUE` (the function computes analytical covariances by default). Note that the computation of analytical covariances can be quite slow for large samples (thousands of observations).

The output of the function `spsurml()` is an 'spsur' object and, consistent with the conventions in the R environment, the `summary()` method prints the output equation by equation.

```
R> summary(spcsur.slm)


Call:
spsurml(formula = spcformula, data = spc, listw = Wspc, type = "slm",
    method = "eigen")
```

```
Spatial SUR model type:  slm

Equation  1
               Estimate Std. Error t value  Pr(>|t|)
(Intercept)_1  1.4947178  0.2467450  6.0577 5.244e-07 ***
UN83_1         0.8053338  0.2558760  3.1474  0.003249 **
NMR83_1       -0.5165301  0.2590369 -1.9940  0.053557 .
SMSA_1        -0.0072526  0.0118566 -0.6117  0.544484
rho_1         -0.5048488  0.2405967 -2.0983  0.042763 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-squared: 0.622
  Equation  2
               Estimate Std. Error t value  Pr(>|t|)
(Intercept)_2  1.7088173  0.2925087  5.8419 1.028e-06 ***
UN80_2        -0.6736472  0.3870209 -1.7406  0.09007 .
NMR80_2        0.7475735  0.3840119  1.9467  0.05919 .
SMSA_2         0.0013487  0.0241871  0.0558  0.95583
rho_2         -0.4816233  0.2557338 -1.8833  0.06754 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-squared: 0.4743


Variance-Covariance Matrix of inter-equation residuals:
  0.0003091646 -0.0003578072
 -0.0003578072  0.0015874436
Correlation Matrix of inter-equation residuals:
  1.0000000 -0.5107461
 -0.5107461  1.0000000

 R-sq. pooled: 0.6601
 Breusch-Pagan: 6.543  p-value: (0.0105)
 LMM: 0.50474  p-value: (0.477)
```

The summary includes standard information for a SUR estimation, disaggregated by equation, such as estimated coefficients, standard deviations, etc. The matrices of covariances and of correlations between the SUR residuals appear next. Finally, `spsurml()` shows some measures of goodness of fit (overall R^2, log-likelihood), the usual Breusch-Pagan (BP) test of diagonality (Breusch and Pagan 1980), and a marginal Lagrange multiplier (LMM) to test for the omitted spatial effects arising from the model that we have estimated (that is, if the type is `slm` or `sdm`, the LMM tests for omitted spatial errors; if the type is `sem` or `sdem`, the LMM tests for omitted spatial lags).

Additional standard methods for **spsur** objects include `anova`, `coef`, `fitted`, `logLik`, `print`, `plot`, `residuals`, and `vcov`. All of them work in the usual way for R standard methods.

Example with `NCOVR`

Next, the `NCOVR` database is used to provide an example of a multiequational Durbin model with a different set of spatially lagged regressors, LU factorization (since the weighting matrix is not symmetric, Cholesky cannot be applied), and numerical covariances (to deal with 12,340 observations):

```
R> ncovrformulaD <-  ~ PS80 + UE80 | PS80 | PS80
R> mlcontrol <- list(fdHess = TRUE, trace = FALSE)
R> ncovrsur.sdm <- spsurml(formula = ncovrformula, data = NCOVR.sf,
+    type = "sdm", Durbin = ncovrformulaD, listw = ncovrlw, method = "LU",
+    control = mlcontrol)
R> summary(ncovrsur.sdm)

Call:
spsurml(formula = ncovrformula, data = NCOVR.sf, listw = ncovrlw,
    type = "sdm", Durbin = ncovrformulaD, method = "LU", control = mlcontrol)


Spatial SUR model type:  sdm

Equation  1
              Estimate Std. Error t value  Pr(>|t|)
(Intercept)_1 2.268071   0.270906  8.3722 < 2.2e-16 ***
PS80_1        1.139491   0.152999  7.4477 1.035e-13 ***
UE80_1        0.185387   0.045721  4.0547 5.061e-05 ***
lag.PS80_1   -1.044969   0.208492 -5.0120 5.486e-07 ***
lag.UE80_1   -0.182819   0.058970 -3.1002   0.00194 **
rho_1         0.670546   0.017641 38.0110 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-squared: 0.3543
  Equation  2
               Estimate Std. Error  t value Pr(>|t|)
(Intercept)_2  0.6826469  0.0460874  14.8120   <2e-16 ***
PS80_2         0.5112242  0.0277382  18.4303   <2e-16 ***
UE80_2         0.0519760  0.0056002   9.2810   <2e-16 ***
SOUTH_2       -0.0321020  0.0355176  -0.9038   0.3661
lag.PS80_2    -0.5477503  0.0377724 -14.5014   <2e-16 ***
rho_2          0.7812021  0.0141816  55.0858   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-squared: 0.5264
  Equation  3
              Estimate Std. Error t value  Pr(>|t|)
(Intercept)_3 1.956636   0.065810  29.731 < 2.2e-16 ***
PS80_3       -1.824512   0.100259 -18.198 < 2.2e-16 ***
lag.PS80_3    1.292710   0.134545   9.608 < 2.2e-16 ***
```

```
rho_3          0.845296   0.011038  76.583 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-squared: 0.6718

Variance-Covariance Matrix of inter-equation residuals:
 31.014989  1.0687653  7.0673167
  1.068765  1.0200470 -0.3444854
  7.067317 -0.3444854 13.3581579
Correlation Matrix of inter-equation residuals:
 1.0000000  0.19001437  0.34721258
 0.1900144  1.00000000 -0.09332274
 0.3472126 -0.09332274  1.00000000

 R-sq. pooled: 0.6293
 Breusch-Pagan: 510.2  p-value: (2.98e-110)
```

A comparison of spatial SUR specifications using likelihood ratios

As is well known ([Elhorst 2014](#)), the spatial models listed in Section 2 are related in a nesting sequence. Given that `spsurml()` operates in an ML framework, we can easily compare two nested models using the corresponding likelihood ratios (LRs) to select the best specification. This is the purpose of the `anova()` method applied to 'spsur' objects. In the case of a Phillips curve, we simply write:

```
R> spcsur.slm <- spsurml(formula = spcformula, data = spc, type = "slm",
+    listw = Wspc, control = list(trace = FALSE))
R> spcsur.sdm <- spsurml(formula = spcformula, data = spc, type = "sdm",
+    listw = Wspc, control = list(trace = FALSE))
R> anova(spcsur.slm, spcsur.sdm)


           logLik df    AIC     BIC LRtest  p.val
model 1: slm 114.10 13 -202.20 -219.19
model 2: sdm 116.78 19 -195.57 -220.40 5.3701 0.4973
```

The output of the function is an 'anova' object showing the value of the corresponding LR test (if the `lrtest` argument is `TRUE` as is the default) and its associated p values, obtained from a χ^2 distribution with degrees of freedom equal to the number of restrictions in the respective null hypothesis. Akaike information criterion (AIC) and log-likelihood (logLik) values for each model are also included. The information provided by the likelihood ratios may help the user define the spatial structure needed for the case under consideration.

3.4. 3SLS based on IV estimation of SUR-SLM and SUR-SDM models

The function `spsur3sls()` estimates spatial SUR models by spatial IVs in a 3SLS framework, as described in Section 2.2.2. The syntax of the function is similar to `spsurml()`.

```
spsur3sls(formula = NULL, data = NULL, na.action, R = NULL, b = NULL,
    listw = NULL, zero.policy = NULL, X = NULL, Y = NULL, G = NULL,
    N = NULL, Tm = NULL, p = NULL, type = "slm", Durbin = NULL,
    maxlagW = NULL, trace = TRUE)
```

The main differences between the two are: (i) the argument `type` now only supports two spatial SUR models, `slm` and `sdm`, and (ii) a new argument is needed, `maxlagW`, to set the maximum order of the spatial lags of the regressors and obtain the matrix of IVs. The default value is `NULL`, which chooses `maxlagW = 2` for SUR-SLM models and `maxlagW = 3` for the SUR-SDM variant.

Example with `NCOVR`

The 3SLS method is especially adequate in cases of large datasets or when the Jacobian matrix is very dense. The `NCOVR` dataset including 3,085 US counties is a good candidate for `spsur3sls()`, for which we simply need:

```
R> ncovrsur.slm.3sls <- spsur3sls(formula = ncovrformula, data = NCOVR.sf,
+    type = "slm", listw = ncovrlw, trace = FALSE)
```

As an alternative to `summary`, the user can apply the `print` method to get an abbreviated version of the results,

```
R> print(ncovrsur.slm.3sls)
```

```
            coeff_1 pval_1 coeff_2 pval_2 coeff_3 pval_3
(Intercept)  9.7711  0.000  2.7205  0.000  7.2895      0
PS80         0.9497  0.000  0.2415  0.000 -1.6132      0
UE80        -0.2083  0.000  0.0891  0.000      NA     NA
SOUTH            NA     NA  0.1603  0.000      NA     NA
rho         -0.2067  0.417  0.2630  0.002  0.4176      0
```

If a variable is not included in an equation, the `print` method shows an `NA` value. As in the case of `spsurml()`, the output of the function `spsur3sls()` is an 'spsur' class object.

3.5. Testing the spatial SUR estimation and re-estimating models

An important feature of **spsur** is its ability to thoroughly test the specification. `spsurml()` routinely reports the BP diagonality test as well as a Lagrange marginal multiplier for omitted spatial effects. To these basic measures, we add three functions, `wald_betas()`, `wald_deltas()`, and `lr_betas()`, which improve specification.

Testing for linear restrictions on coefficients

The function `wald_betas()` is used to test linear restrictions in the β parameters. This function needs three arguments. The first is an 'spsur' class object obtained with the functions `spsurml()` or `spsur3sls()`, the second is a matrix that must reflect the restrictions on the β parameters that we want to test, and the third is a vector with the values of the restrictions under the null hypothesis.

For example, to test whether the intercepts of the two equations of the SUR-SLM model of the Phillips curve (13) are equal:

$$H_0 : \beta_{10} = \beta_{20} \quad \text{vs} \quad H_A : \beta_{10} \neq \beta_{20}$$

It is necessary to write the code,

```
R> R1 <- matrix(c(1, 0, 0, 0, -1, 0, 0, 0), nrow = 1)
R> b1 <- matrix(0, ncol = 1)
R> wald_betas(spcsur.slm, R = R1, b = b1)

    Wald test on beta parameters

data:  spc
Wald test = 0.39767, df = 1, p-value = 0.5283
```

The output of the test is an 'htest' object. In this case, the two intercepts are not statistically different from each other, and the correct specification must therefore include the same intercept in both equations. The constrained version of the spatial Phillips curve is:

$$\begin{aligned}
WAGE_{83} &= \beta_0 + \beta_{11}\,UN_{83} + \beta_{12}\,NMR_{83} + \beta_{13}\,SMSA + \lambda_1 WAGE_{83} + \varepsilon_{83}\\
WAGE_{81} &= \beta_0 + \beta_{21}\,UN_{80} + \beta_{22}\,NMR_{80} + \beta_{23}\,SMSA + \lambda_2 WAGE_{81} + \varepsilon_{81}
\end{aligned} \quad (15)$$

Both functions, `spsurml()` and `spsur3sls()`, allow for a restricted ML or 3SLS estimation of the SUR model, using the arguments `R = R1` and `b = b1`, which expedites the estimation problem. The output of the constrained ML estimation appears below.

```
R> R1 <- matrix(c(1, 0, 0, 0, -1, 0, 0, 0), nrow = 1)
R> b1 <- matrix(0, ncol = 1)
R> spcsur.slm.restricted <- spsurml(formula = spcformula, data = spc,
+    type = "slm", listw = Wspc, R = R1, b = b1,
+    control = list(trace = FALSE))
R> print(spcsur.slm.restricted)
```

	coeff_1	pval_1	coeff_2	pval_2
(Intercept)	1.5786	0.000	1.5786	0.000
UN83	0.7946	0.003	NA	NA
NMR83	-0.4873	0.064	NA	NA
SMSA	-0.0056	0.632	0.0029	0.906
UN80	NA	NA	-0.6193	0.120
NMR80	NA	NA	0.7083	0.079
rho	-0.5856	0.006	-0.3705	0.052

Testing for linear restrictions on the spatial parameters

Similarly, the purpose of the function `wald_deltas()` is to obtain a Wald test to evaluate linear restrictions on the spatial parameters of the model (λ or ρ parameters). The use of the function is totally analogous to `wald_betas()` and extends to the ML and 3SLS methods. For

example, the user may be interested in testing whether the λ parameters of the two equations of the constrained version of the previous Phillips curve are equal; that is:

$$H_0 : \rho_1 = \rho_2 \quad \text{vs} \quad H_A : \rho_1 \neq \rho_2 \tag{16}$$

```
R> R2 <- matrix(c(1,-1), nrow = 1)
R> b2 <- matrix(0, ncol = 1)
R> wald_deltas(spcsur.slm.restricted, R = R2, b = b2)

    Wald test on spatial delta parameters

data:  spc
Wald test = 16.932, df = 1, p-value = 3.874e-05
```

In this case, the Wald test clearly rejects the hypothesis of (16), confirming that the specification (15) with two different parameters of spatial dependence is correct.

3.6. Spatial impacts in the SUR framework

spsur includes the function `impactspsur()` to evaluate spatial effects. The function is a wrapper in a multiequational context of the `impacts()` function available in[3] **spatialreg** (see Bivand and Piras 2015, for details). As a consequence, the arguments of `impactspsur()` are a subset of the corresponding arguments in `impacts()`. Specifically, the full syntax of `impactspsur()` is:

```
impactspsur(obj, ..., tr = NULL,R = NULL, listw = NULL, evalues = NULL,
    tol = 1e-06, empirical = FALSE, Q = NULL)
```

The argument `obj` includes an 'spsur' object (the output of `spsurml()`, `spsur3sls()` or `spsurtime()`). The rest of the arguments can be checked using either the help of `impacts()` or the help of `impactspsur()`.

Example with `spc`

Below, an example of how to compute the impacts using the previous Durbin model estimated for `NCOVR.sf` data is shown. We also compute the vector of traces of powers of the spatial weights matrix (using the `trW()` function of **spatialreg**) to speed up the computations.

```
R> ncovrw <- as(spdep::listw2mat(ncovrlw), "CsparseMatrix")
R> ncovrtrw <- spatialreg::trW(ncovrw, type = "MC")
R> ncovrsur.sdm.impacts <- impactspsur(ncovrsur.sdm, tr = ncovrtrw, R = 1000)
```

---

[3]Recently Kelejian and Piras (2020) have developed a formula to compute the variance for spatial impacts in two contexts: when there is no endogenous regressor in each equation (apart from the spatial lag) and when there are endogenous regressors and the general system of equations is not fully specified. We believe that in our spatial SUR contexts, the first approach is quite promising because it would allow users to make inferences about spatial impacts without resorting to simulations. In fact, it can also be applied when the distribution of the error term is nonparametric but heteroskedastic and autocorrelated consistent covariances are available from the estimation process. On the other hand, the second approach (allowing for endogenous regressors in a multiequational context) is especially interesting for spatial simultaneous equation systems and, to our knowledge, it is best suited for the **spse** package, which is specialized in this type of model (see Piras 2018, for details). We will leave these promising lines of research for future updates of the **spsur** package.

For model types `"slm"`, `"sdm"`, `"sarar"`, or `"gnm"`, the output of the function is a list of ‘`lagImpact`’ objects, including the impacts for each equation of the model. Each element of the list can be printed in the usual way:

```
R> summary(ncovrsur.sdm.impacts[[1]], zstats = TRUE, short = TRUE)
```

```
Impact measures (mixed, trace):
          Direct    Indirect        Total
PS80_1 1.1090557 -0.8221535 0.286902202
UE80_1 0.1790469 -0.1712532 0.007793673
===========================================================
Simulation results ( variance matrix):
===========================================================
Simulated standard errors
            Direct   Indirect      Total
PS80_1 0.14804029 0.4662831 0.4532275
UE80_1 0.04400361 0.1204696 0.1137512

Simulated z-values:
         Direct  Indirect       Total
PS80_1 7.535948 -1.805916 0.60357246
UE80_1 4.065629 -1.417654 0.07136721

Simulated p-values:
        Direct    Indirect Total
PS80_1 4.8406e-14 0.070931 0.54613
UE80_1 4.7903e-05 0.156292 0.94311
```

Users can obtain the the impacts for the second equation in the same way,

```
R> summary(ncovrsur.sdm.impacts[[2]], zstats = TRUE, short = TRUE)
```

The `impactspsur()` function also allows us to estimate impacts of the `slx` or `sdem` models. These models do not include a spatial lag of the dependent variable, so the impacts can be directly computed from the estimated coefficients (no simulation is needed). To compute the variances for the impacts, we use the function `estimable()` from the **gmodels** package. The output of `impactspsur()` for models `"slx"` or `"sdem"` is a list of ‘`mixedImps`’ objects (one for each equation) and can be printed in the usual way. Below, an example is provided for `NCOVR.sf` data:

```
R> ncovrsur.slx <- spsurml(formula = ncovrformula, data = NCOVR.sf,
+    type = "slx", listw = ncovrlw, control = list(trace = FALSE))
R> ncovrsur.slx.impacts <- impactspsur(ncovrsur.slx)
```

The output is omitted to save space.

4. Simulated spatial SUR datasets

This section describes an additional function included in **spsur** that offers the possibility of generating random datasets with the required dimensions and spatial structure. This function, `dgp_spsur()`, has great flexibility and may be of particular interest in two cases: (i) as part of a larger simulation where the users need random datasets with specific features (i.e., for their own research projects) or (ii) with the aim of showing specific properties of spatial SUR models and that inferential procedures related to them (i.e., for teaching activities).

`dgp_spsur()` is a data generating process (DGP) where the user decides the dimensions of the required dataset, specifies the values of the parameters that intervene in the corresponding SUR model, and selects the distribution functions from which the regressors and the random terms are to be drawn. The syntax of the function is the following:

```
dgp_spsur(Sigma, Tm = 1, G, N, Betas, Thetas = NULL, rho = NULL,
    lambda = NULL, p = NULL, listw = NULL, X = NULL,
    type = "matrix", pdfU = "nvrnorm", pdfX = "nvrnorm")
```

The dimensions of the dataset are defined by the arguments Tm, N, and G. The argument `Sigma` specifies the covariance matrix among the residuals of the G equations, which is the core of the SUR model. Of course, a $(N \times N)$ spatial weighting matrix should be uploaded in the argument `listw`.

A fundamental piece of information is the argument p, which defines the number of regressors (**X** variables) that appear in each equation. If p is a scalar, every equation has the same number of regressors, whereas a $1 \times G$ vector indicates that the G equations contain a different number of regressors. Subsequently, the user should specify the parameters that intervene in the equations of the SUR. This is the purpose of the arguments `Betas`, `Thetas`, `rho`, and `lambda`, which are defined as row vectors of the adequate dimensions. `Betas` is a row vector of the order $1 \times K$, where K is equal to pG, or to $K = \Sigma_{j=1}^{G} p_j$ if p is a row vector. The values assigned to these parameters determine the type of spatial model required. For example, if `Thetas = NULL`, `rho = NULL` but `lambda` and `Betas` are not NULL, this means that the user is specifying a SUR-SEM model, or a SUR-SDEM model in the case of `rho = NULL` but `Thetas`, `Betas`, and `lambda` are different from NULL.

There are two possibilities to build the matrix **X** of regressors. In Monte Carlo experiments, it is usual to maintain the values of the regressors fixed and draw random matrices only for the error terms. If this is the case, the user should upload the required **X** matrix in the argument X, which must be consistent with the dimensions of the SUR model. If `X = NULL`, the user randomly draws this matrix using the function `dgp_spsur()`. This is the purpose of the argument `pdfX`, which uses a multivariate standard normal distribution. The alternative is a Uniform $[0,1]$ distribution for each regressor. In both cases, the regressors are generated independently. Finally, the argument `dfU` selects the multivariate probability distribution function to draw the values of the error terms.

Output types can be selected using the `type` argument. By defect, the output of `dgp_spsur()` is a vector called **Y** of the order $(NTmG \times 1)$, with the values generated for the explained variable in the G equations of the spatial SUR model. If the argument X is set to NULL, the user will receive another matrix called **X** with the values generated for the regressors of the SUR.

4.1. A toy example. How to create random datasets

In this section, we present a toy example of a spatial SUR process with two equations for 537 individuals defined on a regular hexagonal lattice. The geometry of the lattice is defined on a 1×1 square using the **sf** package.

```
R> sfc <- st_sfc(st_polygon(list(rbind(c(0, 0), c(1, 0), c(1, 1), c(0, 1),
+    c(0, 0)))))
R> hexs <- st_make_grid(sfc, cellsize = 0.049, square = FALSE)
R> hexs.sf <- st_sf(hexs)
```

The weight matrix that defines the neighborhood structure, **W**, is obtained using the rook criteria (common border).

```
R> W <- nb2listw(poly2nb(as(hexs.sf, "Spatial"), queen = FALSE))
```

For the DGP example, we consider two equations and 2 regressors, plus the intercepts. We generate a SUR-SLM model with different levels of spatial dependence in each equation, as shown in (17)).

$$
\begin{aligned}
Y_1 &= 1 + 0.2\ \text{W}Y_1 + 2\ X_{11} + 3\ X_{12} + \epsilon_1 \\
Y_2 &= 1 + 0.8\ \text{W}Y_2 -\ X_{21} + 0.5\ X_{22} + \epsilon_2 \\
\text{cor}(\epsilon_1, \epsilon_2) &= 0.5
\end{aligned}
\tag{17}
$$

The function `dgp_spsur()` produces the simulated data for the two variables using this simple set of arguments

```
R> Sigma <- matrix(c(1,0.5,0.5,1),nrow = 2)
R> Betas <- c(1,2,3,2,-1,0.5)
R> rho <- c(0.2,0.8)
R> set.seed(123)
R> dgp.spatial.sur <- dgp_spsur(Sigma = Sigma, Betas = Betas, Thetas = NULL,
+    lambda = NULL, rho = rho, Tm = 1, G = 2, N = 537, p = 3, listw = W)
```

The output of this function is a list with the data, dependent variables, and regressors in matrix forms. Next, we estimate a SUR-SIM model using the method `residuals` to obtain the residuals.

```
R> sur.dgp.sim <- spsurml(Y = dgp.spatial.sur$Y, X = dgp.spatial.sur$X,
+    type = "sim", G = 2, Tm = 1, N = 537, p = 3, listw = W,
+    control = list(trace = FALSE))
R> res <- residuals(sur.dgp.sim)
```

Figure 3 plots them. Note that the residuals of the second equation in (17) show a stronger spatial structure. Finally, we estimate the *true* model, a SUR-SLM model, as usual,

```
R> sur.dgp.slm <- spsurml(Y = dgp.spatial.sur$Y, X = dgp.spatial.sur$X,
+    type = "slm", G = 2, Tm = 1, N = 537, p = 3, listw = W,
+    control = list(trace = FALSE))
R> summary(sur.dgp.slm)
```
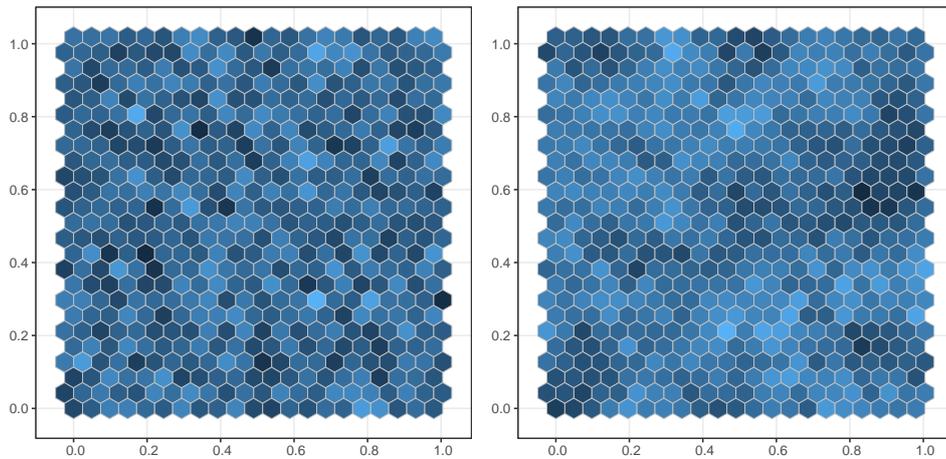
Figure 3: Spatial pattern of the SUR-SIM residuals, ϵ_1 (left) ϵ_2 (right).

```
Call:
spsurml(listw = W, type = "slm", X = dgp.spatial.sur$X, Y = dgp.spatial.sur$Y,
    G = 2, N = 537, Tm = 1, p = 3, control = list(trace = FALSE))

Spatial SUR model type:  slm

Equation  1
           Estimate Std. Error t value  Pr(>|t|)
Intercep_1 1.050658   0.052887  19.866 < 2.2e-16 ***
X1_1       2.009105   0.038119  52.706 < 2.2e-16 ***
X1_2       2.984680   0.035863  83.225 < 2.2e-16 ***
rho_1      0.161010   0.021984   7.324 4.746e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-squared: 0.9292
  Equation  2
            Estimate Std. Error  t value  Pr(>|t|)
Intercep_2  2.066812   0.219397   9.4204 < 2.2e-16 ***
X2_1       -0.978890   0.038798 -25.2304 < 2.2e-16 ***
X2_2        0.469211   0.036570  12.8305 < 2.2e-16 ***
rho_2       0.791277   0.021862  36.1946 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-squared: 0.8029

Variance-Covariance Matrix of inter-equation residuals:
 0.9994753 0.5283877
 0.5283877 1.0096595
Correlation Matrix of inter-equation residuals:
 1.0000000 0.5259921
 0.5259921 1.0000000
```

```
R-sq. pooled: 0.9632
Breusch-Pagan: 148.6  p-value: (3.56e-34)
LMM: 6.4689  p-value: (0.011)
```

As expected, the results are quite satisfactory. All the estimated parameters are within the interval (theoretical value plus/minus two standard deviations). The Breusch-Pagan test (with a value of 148.57) clearly rejects the assumption of diagonality. The correlation obtained for the SUR equations are slightly inflated (0.516 on average for an expected value of 0.5), whereas the marginal multiplier (6.47) clearly confirms the adequacy of the spatial lag model.

5. spsur application to mobility due to COVID-19

There is broad consensus that this pandemic has been one of the most relevant events that has occurred during recent decades with impacts (social, medical, economic, etc.) that will last for a long time. However, in spite of the dramatic consequences, social distancing and the reduction of individual mobility are seen as the most effective strategies to slow down the spread of the disease.

This section illustrates the practical use of the **spsur** package to develop a spatial SUR model and explain the effect of COVID-19 on two types of individual mobility (controlling for several factors). In particular, we focus on the case of Spain, using data at a provincial level (NUTS3 in Eurostat terminology) during the first weeks of the pandemic[4]. The hypotheses that we explore is that the high incidence of COVID-19 dissuades individuals from leaving their homes and therefore reduces mobility.

For this illustration, we consider two different indicators of individual mobility: "Within" and "Exits". The first is defined by the ratio between the number of intra-provincial trips within province "i" in week "t" compared to a pre-COVID week (February, 14–21, 2020). The second is defined in the same way but for inter-provincial trips (see Table 1). Figure 4 shows the weekly evolution of these mobility indices; Figures 4A for intra-provincial and 4B for inter-provincial trips. Note the sharp reduction of the mobility indices, both intra and inter, occurring in week 4. This is not accidental, since week 4 is the beginning of the state of emergency (StEm) declared by the Spanish Government. Weeks 6 and 7 (in orange), correspond to the tightening of restrictive measures when no essential economic activities were allowed. The StEm remained in effect until June 21, the end of week 17.

Several variables could have impact on both mobility indices (Table 1). We consider three categories of variables: government restriction orders, socio-economic characteristics, and

---

[4]In response to the rapidly growing number of COVID-19 cases, the Spanish Government (like many others) enacted orders to reduce individual mobility, declaring a state of emergency (StEm) on March 14, 2020. In practice, the StEm implied a lockdown of the population. Many non-essential service activities, such as education, leisure and entertainment, commerce (except for food, health or electronics) were suspended. On March 28, 2020, the lockdown also reached the sectors of manufacturing and construction, with grave implications for the Spanish economy. On April 10, 2020, both sectors were allowed to resume their activities, as there was a clear indication that the lockdown had been successful, and on June 23, the StEm was lifted. During the same time, the Spanish Government initiated a big data project with the purpose of collect information about the mobility of the population by monitoring the location of all the cell phones in national territory (in fact, 80% of mobile phones are under control). The daily collection began on February 14 and includes all movements exceeding 500 meters. This information also distinguishes short range or intra-provincial trips whose origin and destination are both inside the same province, inter-provincial, or long range trips, where the province of origin differs from that of the destination.
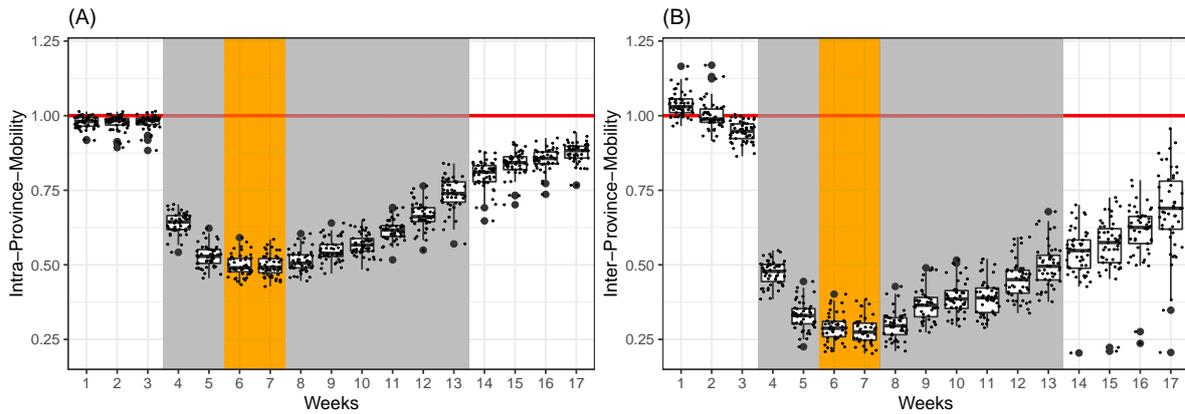
Figure 4: (A) Weekly intra-province changes in mobility compared to pre-COVID-19 week (February 14–20); (B) Weekly inter-province changes in mobility compared to pre-COVID-19 week (February 14–20). In orange, total lockdown, weeks 6 and 7 (from March 28 to April 10). The red line marks the mobility level = 1 (the reference week).

COVID-19 incidence. With respect to government restriction orders, the Spanish Government declared a partial lockdown with stay-at-home orders for citizens for 10 weeks (weeks 4–13), but essential economic activities were permitted. In the middle of this period (two weeks), the lockdown was total, and essential economic activities were not permitted. Therefore, it is expected that during this period, mobility would decrease. Therefore, the expected sign of these variables explaining mobility is negative.

Even though mobility restrictions were dictated for the entire country, the Government's orders were unevenly followed in different provinces. Several factors related with the socio-economic characteristics can explain these regional differences.

In the first place, differences in population density have an impact on mobility (Brodeur, Grigoryeva, and Kattan 2021). In this case, high levels of population density (the provinces of Madrid or Barcelona) can change individual behavior due to the fact that the virus spreads mainly through interpersonal interactions (expected negative sign). In the second place, something similar occurs in provinces with a high proportion of older people, whose ability to move is more restricted (Brodeur *et al.* 2021).They are also more responsive to restriction orders (Páez *et al.* 2021, Engle, Stromme, and Zhou (2020)). In the third place, provinces with high levels of essential services reduce mobility less than provinces with low levels of essential services (expected negative sign, Abdullah, Dias, Muley, and Shahin 2020). Lastly, the incidence of COVID-19 could influence movement in the sense that evidence of the disease dissuades individuals from leaving their homes, and therefore mobility is reduced (Engle *et al.* 2020, Abdullah *et al.* (2020)). Table 1 summarizes all the variables considered in this illustration and their expected signs.

Figure 5A represents the weekly incidence of the pandemic (log of cases detected per 100,000 inhabitants). The incidence has an inverted-V shape that peaks on week 6. Looking at this figure, it is apparent that the incidence of COVID-19 has a strong impact on mobility, since movement decreases considerably as the virus intensifies. Figure 5B shows the spatial distribution of COVID-19. A strong spatial pattern emerges, confirming the spatial diffusion of the pandemic and therefore, spatial effects in the model are expected.
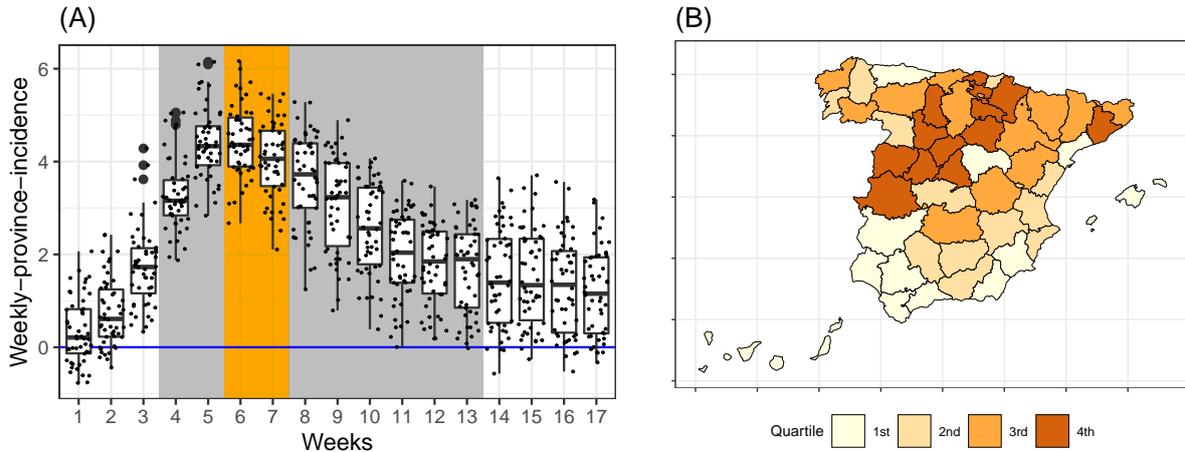
Figure 5: (A) Weekly incidence, total positive PCR cases per 100,000 inhabitants (in logs). In gray, partial lockdown weeks 4–5 and 8–13 (from March 14–27 and April 11–23). In orange, total lockdown, weeks 6 and 7 (from March 28 to April 10). The blue line marks zero incidence. (B) Quartiles of provincial COVID-19 incidence for week 6 (in log). Similar spatial pattern in other weeks.

Variable	Sort Description	Sign
Within	Number of trips in week "t" within province "i" (compared to the pre-COVID reference week)	
Exit	Number of trips in week "t" with origin in province "i" and arrival to another (compared to the pre-COVID reference week)	
Emergence	Dummy variable. 1 if StEm is active in week "t". Essential activities (e.g. food, health) are allowed. 0 otherwise	$(-)$
EmergenceTotal	Dummy variable. 1 if StEm is active in week "t". Essential activities are not allowed. 0 otherwise	$(-)$
Density	Inhabitants (in thousands) per km2 in province "i"	$(-)$
Old65	Percentage of population aged 65 and older in province "i"	$(-)$
Essential	Percentage of firms in province "i" with essential activities	$(+)$
Incidence	Weekly incidence in the week "$t-1$" in the province "i" (in logs)	$(-)$

Table 1: List of dependent and independent variables for the spatial SUR model and expected sign. See the help facility in the **spsur** package for details and formal definitions.

5.1. The SUR approach

The dataset with the two mobility indices and the control variables for the 50 spatial units (provinces) and 17 temporal periods (weeks from February 14 to June 21, 2020) is available in the **spsur** package,

```
R> data("spain.covid", package = "spsur")
```

A baseline SUR-SIM model with $G = 2$ equations (for inter and intra movements, respec-

tively), which does not include spatial effects can be specified as follows:

$$\text{Within}_{i,t} = \beta_{10} + \beta_{11}\text{Emergence}_t + \beta_{12}\text{EmergenceTotal}_t + \beta_{13}\text{Density}_i + \beta_{14}\text{Old65}_i +$$
$$\beta_{15}\text{Essential}_i + \beta_{16}\text{Incidence}_{i,t-1} + \epsilon_{i,t}^W$$

$$\text{Exits}_{i,t} = \beta_{20} + \beta_{21}\text{Emergence}_t + \beta_{22}\text{EmergenceTotal}_t + \beta_{23}\text{Density}_i + \beta_{24}\text{Old65}_i +$$
$$\beta_{25}\text{Essential}_i + \beta_{26}\text{Incidence}_{i,t-1} + \epsilon_{i,t}^E$$

$$\text{cor}(\epsilon_{i,t}^W, \epsilon_{j,t}^E) = \sigma_{i,j} ; \quad t = 1, \dots, 17; \quad i = 1, \dots, 50$$

$$(18)$$

To estimate this model using the **spsur** package, first, the two equations of model (18) can be specified using the **Formula** package. In this case, only a single sequence of explanatory variables is included on the right-hand side of the formula because these factors are the same for both equations.

```
R> formula <- Within | Exits ~ Emergence + EmergenceTotal + Density +
+    Old65 + Essential + Incidence
```

This specification can be estimated for maximum likelihood with the function `spsurml()`

```
R> Tm <- 17
R> covid.sim <- spsurml(formula = formula, data = spain.covid, type = "sim",
+    Tm = Tm, control = list(trace = FALSE))
R> summary(covid.sim)


Call:
spsurml(formula = formula, data = spain.covid, type = "sim",
    Tm = Tm, control = list(trace = FALSE))


Spatial SUR model type:  sim

Equation  1
                  Estimate  Std. Error   t value   Pr(>|t|)
(Intercept)_1    0.75220735  0.04712812   15.9609  < 2.2e-16 ***
Emergence_1     -0.24061703  0.00593652  -40.5316  < 2.2e-16 ***
EmergenceTotal_1 -0.05211290  0.00795879   -6.5478  7.731e-11 ***
Density_1       -0.04233591  0.01700889   -2.4890   0.012905 *
Old65_1          0.34461164  0.06352238    5.4250  6.636e-08 ***
Essential_1      0.00177808  0.00066556    2.6716   0.007623 **
Incidence_1     -0.03576960  0.00219800  -16.2737  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-squared: 0.8643
  Equation  2
                  Estimate Std. Error   t value   Pr(>|t|)
(Intercept)_2    0.4349503  0.1135108    3.8318 0.0001319 ***
```

```
Emergence_2      -0.3132801  0.0142985 -21.9101 < 2.2e-16 ***
EmergenceTotal_2 -0.0585423  0.0191692  -3.0540 0.0022937 **
Density_2        -0.0513039  0.0409669  -1.2523 0.2106254
Old65_2           0.6620623  0.1529974   4.3273 1.598e-05 ***
Essential_2       0.0035088  0.0016030   2.1888 0.0287457 *
Incidence_2      -0.0405096  0.0052940  -7.6519 3.312e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-squared: 0.6338

Variance-Covariance Matrix of inter-equation residuals:
 0.004261810 0.007104235
 0.007104235 0.024723421
Correlation Matrix of inter-equation residuals:
 1.0000000 0.6920954
 0.6920954 1.0000000


 R-sq. pooled: 0.7505
 Breusch-Pagan: 407.1  p-value: (1.53e-90)
```

The residuals of both equations are highly correlated (0.69), and the Breush-Pagan test (407.1) confirms the adequacy of the SUR specification. All the variables, with the exception of `Density` in the second equation, are significant. The signs of the two dummy variables, (`Emergence` and `EmergenceTotal`) are negative, as expected. Population density (Density) is significant and negative in the first equation and not significant in the second one. The variable for older people, `Old65`, contrary to expectations, has a positive sign, which is difficult to justify in this first model. The percentage of essential services (`Essential`) has a positive impact on movement in the sense that greater economic activity promotes mobility. Finally, a strong incidence of the disease in previous weeks, `Incidence`, has a negative sign, showing that a high incidence of COVID-19 discourages individuals from moving.

Given the framework that supports our research and using Spanish provinces as observation units, it is advisable to test for the presence of (omitted) spatial effects, especially spatial autocorrelation. We consider a weighting matrix based on the queen criteria to test the hypothesis of spatial independence in the residuals. The **spdep** package can be used at this moment. Note that there are three provinces with no neighbors (in the Balearic and Canary islands), so the option `zero.policy = TRUE` will be necessary, as appears in the code below.

```
R> listw <- spdep::nb2listw(listw, style = "W", zero.policy = TRUE)
```

The **spsur** function `lmtestspsur()` helps test the SUR-SIM model against any other specification through the corresponding Lagrange multipliers. The results, as shown below, detect a strong spatial structure that has been omitted in the SUR-SIM: all the LM tests, robust and non-robust, reject their corresponding null hypotheses.

```
R> covid.lmtest <- lmtestspsur(formula = formula, data = spain.covid,
+    Tm = Tm, listw = listw, zero.policy = TRUE)
R> pr.covid.lmtest <- sapply(covid.lmtest, broom::tidy)
R> print(as.data.frame(t(pr.covid.lmtest)))
```
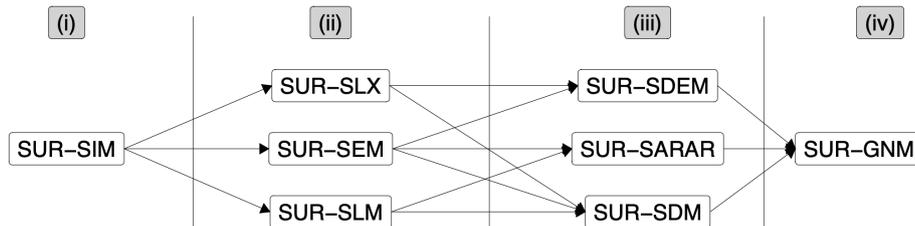
Figure 6: Nesting structure for the spatial models. The model to the left of the arrow is nested in the model to the right of the arrow.

```
  statistic       p.value parameter       method
1  547.8406 1.091246e-119         2    LM-SUR-SLM
2  1187.091 1.684355e-258         2    LM-SUR-SEM
3  143.2781  7.718816e-32         2   LM*-SUR-SLM
4  782.5287 1.191412e-170         2   LM*-SUR-SEM
5  1330.369 8.661217e-287         4  LM-SUR-SARAR
```

5.2. Spatial SUR models. Selection strategy.

The next question is obvious: if the SUR-SIM model is not appropriate, which is the best specification, from a spatial perspective, to model the relation between mobility and COVID-19? Since the LM tests don't give an a priori specification preference, we can estimate all the possible spatial SUR models listed in Figure 6 and obtain several statistical measures – LogLik, AIC, BIC (Bayesian information criterion) – to select the most suitable model.

The **spsur** user can estimate all the models by changing the `type` argument. For SUR-SLX, SUR-SDM, SUR-SDEM, and SUR-GNM specifications, we consider only spatial lags for a subset of the independent variables. The `Durbin` option can be used for this purpose.

```
R> formulaD <- ~ Essential + Incidence
R> covid.slx <- spsurml(formula = formula, data = spain.covid, type = "slx",
+    Tm = Tm, Durbin = formulaD, listw = listw, zero.policy = TRUE)
R> covid.slm <- spsurml(formula = formula, data = spain.covid, type = "slm",
+    Tm = Tm, listw = listw, zero.policy = TRUE)
R> covid.sem <- spsurml(formula = formula, data = spain.covid, type = "sem",
+    Tm = Tm, listw = listw, zero.policy = TRUE)
R> covid.sdm <- spsurml(formula = formula, data = spain.covid, type = "sdm",
+    Tm = Tm, Durbin = formulaD, listw = listw, zero.policy = TRUE)
R> covid.sdem <- spsurml(formula = formula, data = spain.covid, type = "sdem",
+    Tm = Tm, Durbin = formulaD, listw = listw, zero.policy = TRUE)
R> covid.sarar <- spsurml(formula = formula, data = spain.covid,
+    type = "sarar", Tm = Tm, listw = listw, zero.policy = TRUE)
R> covid.gnm <- spsurml(formula = formula, data = spain.covid, type = "gnm",
+    Tm = Tm, Durbin = formulaD, listw = listw, zero.policy = TRUE)
```

At this point, the `anova` method can be applied to objects of the 'spsur' class to get information about the LogLik of each model and the information criteria AIC and BIC. Note that

the `lrtest = FALSE` must be selected because the eight models are not nested.

```
R> anova(covid.sim, covid.slx,covid.slm, covid.sem,covid.sdm, covid.sdem,
+    covid.sarar, covid.gnm, lrtest = FALSE)
```

```
               logLik df     AIC     BIC
model 1: sim    1758.1 17 -3482.2 -3504.4
model 2: slx    1851.0 21 -3660.1 -3687.5
model 3: slm    2076.9 19 -4115.8 -4140.6
model 4: sem    2437.8 19 -4837.6 -4862.5
model 5: sdm    2297.6 23 -4549.1 -4579.2
model 6: sdem   2487.0 23 -4927.9 -4958.0
model 7: sarar  2472.6 21 -4903.3 -4930.7
model 8: gnm    2491.8 25 -4933.7 -4966.3
```

The results show a clear decrease in both AIC and BIC for the most complex models. Moreover, models SUR-SDEM and SUR-GNM show the best performance in the information criteria. In cross-section, the criteria of AIC and BIC have been used to select the correct specification (Agiakloglou and Tsimpanos 2021). Given that both models are nested (see Figure 6), we can use the LR test to choose between them. Again, the `anova()` method can be used, but in this case with `lrtest = TRUE` to show the LR test of both nested models.

```
R> anova(covid.sdem, covid.gnm, lrtest = TRUE)
```

```
               logLik df     AIC     BIC LRtest     p.val
model 1: sdem   2487.0 23 -4927.9 -4958.0
model 2: gnm    2491.8 25 -4933.7 -4966.3 9.7482 0.007642
```

At this point, it is necessary to note that although in terms of likelihood, the SUR-GNM is a higher model than SUR-SDEM, the SUR-GNM could lead to an overestimated model and to problems identifying spatial dependence parameters (e.g. Burridge, Elhorst, and Zigova 2016; Vega and Elhorst 2015). Therefore, the alternative SUR-SDEM is finally selected. For example, for spatial cross-section models Elhorst (2014, page 33) points out his preference for SDEM versus GNM. In contrast, Vega and Elhorst (2015, page 341) claim that "only the parameters of a spatial econometric model with all possible spatial interaction effects based on arbitrary spatial weights matrices have not been proved to be free of this type of identification problem". Finally, we decide to follow the specific-to-general strategy and consider the SUR-GNM as the final model.

The `print()` method showing the coefficients and *p* values of this specification follows.

```
R> print(covid.gnm)
```

```
              coeff_1 pval_1 coeff_2 pval_2
(Intercept)    0.9016  0.000  0.4831  0.000
Emergence     -0.2858  0.000 -0.3621  0.000
EmergenceTotal -0.0846  0.000 -0.0592  0.043
Density       -0.1045  0.000 -0.0758  0.000
```

```
Old65          0.0165  0.784 -0.3266  0.013
Essential      0.0005  0.222  0.0021  0.015
Incidence     -0.0160  0.000 -0.0108  0.017
lag.Essential  0.0010  0.011  0.0025  0.000
lag.Incidence -0.0211  0.000 -0.0069  0.561
rho           -0.0827  0.024  0.1086  0.016
lambda         0.8242  0.000  0.7964  0.000
```

The parameters of spatial error dependence λ_1, λ_2 take values close to 1 (the maximum value) and are highly significant as a symptom of the strong spatial structure in the residuals of this model, probably due to the omission of relevant variables.

On the other hand, the parameters of substantive spatial dependence ρ_1, ρ_2 take values close to zero but are significant, with p values lower than 0.05. In the first equation, the ρ_1 parameter has an opposite sign to λ_1. This might be a symptom of the erroneous identification of the spatial dependence parameters. However, the signs of ρ_2 and λ_2 are both positive in the second equation. Again, the possibility of selecting the SUR-SDEM specification as an alternative to SUR-GNM is on the table.

5.3. Testing linear restrictions on the parameters of the model

Several tests can be used to reduce the model specification and simplify the model. In the first place, it is possible to test whether government restrictions have had the same impact on reducing mobility when people stay inside and when they leave the province. The function `wald_betas()` can be used to obtain the Wald test of coefficient equality. Following the notation of (18), the null hypothesis is,

$$H_0 : \beta_{11} = \beta_{21} \quad \text{vs} \quad H_A : \beta_{11} \neq \beta_{21}$$

The code to test this hypothesis is (see `?spsur::wald_betas`)

```
R> R <- matrix(0, ncol = Tm + 1, nrow = 1)
R> R[2] <- 1
R> R[11] <- -1
R> wald_betas(covid.gnm, R = R, b = 0)

    Wald test on beta parameters

data:  spain.covid
Wald test = 16.79, df = 1, p-value = 4.174e-05
```

The test for the second restriction is the following,

$$H_0 : \beta_{12} = \beta_{22} \quad \text{vs} \quad H_A : \beta_{12} \neq \beta_{22} \tag{19}$$

The code to test the hypothesis (19) is,

```
R> R <- matrix(0, ncol = Tm + 1 , nrow = 1)
R> R[3] <- 1
R> R[12] <- -1
R> wald_betas(covid.gnm, R = R, b = 0)
```

```
    Wald test on beta parameters

data:  spain.covid
Wald test = 0.97674, df = 1, p-value = 0.323
```

Based on the previous results, the model can be estimated by imposing the same coefficient for `EmergenceTotal` in both equations. The user can obtain the restricted model including the arguments R = R and b = 0 in the `spsurml()` function,

```
R> R <- matrix(0, ncol = Tm + 1, nrow = 1)
R> R[3]  <- 1
R> R[12] <- -1
R> covid.gnm.r <- spsur::spsurml(formula = formula, data = spain.covid,
+    type = "gnm", Tm = Tm, Durbin = formulaD, R = R, b = 0, listw = listw,
+    zero.policy = TRUE, control = list(trace = FALSE))
R> print(covid.gnm.r)
```

```
                coeff_1 pval_1 coeff_2 pval_2
(Intercept)      0.9017  0.000  0.4850  0.000
Emergence       -0.2857  0.000 -0.3588  0.000
EmergenceTotal  -0.0847  0.000 -0.0847  0.000
Density         -0.1046  0.000 -0.0760  0.000
Old65            0.0166  0.783 -0.3325  0.011
Essential        0.0005  0.225  0.0021  0.016
Incidence       -0.0160  0.000 -0.0100  0.025
lag.Essential    0.0010  0.013  0.0025  0.000
lag.Incidence   -0.0209  0.000 -0.0069  0.562
rho             -0.0806  0.027  0.1062  0.018
lambda           0.8234  0.000  0.7963  0.000
```

In this model, the variable `Old65` in the first equation, and `lag.Incidence` in the second one can be removed because they are not significant. To obtain a new model excluding these variables, we can rewrite the formula with different variables on the right-hand side of the `formula`. The R vector must also be adapted to the new dimension of the model,

```
R> formula <- Within | Exits ~
+    Emergence + EmergenceTotal + Density + Essential + Incidence |
+    Emergence + EmergenceTotal + Density + Old65 + Essential + Incidence
R> formulaD <- ~ Essential + Incidence | Essential
R> R <- matrix(0, ncol = 16 , nrow = 1)
R> R[3]  <- 1
R> R[11] <- -1
R> covid.gnm.rr <- spsurml(formula = formula, data = spain.covid,
+    type = "gnm", Tm = Tm, Durbin = formulaD, R = R, b = 0, listw = listw,
+    zero.policy = TRUE, control = list(trace = FALSE))
R> print(covid.gnm.rr)
```

```
              coeff_1 pval_1 coeff_2 pval_2
(Intercept)    0.9003  0.000  0.4749  0.000
Emergence     -0.2857  0.000 -0.3585  0.000
EmergenceTotal -0.0849 0.000 -0.0849  0.000
Density       -0.1046  0.000 -0.0744  0.000
Essential      0.0005  0.145  0.0022  0.006
Incidence     -0.0158  0.000 -0.0098  0.027
lag.Essential  0.0010  0.013  0.0024  0.000
lag.Incidence -0.0199  0.000     NA     NA
Old65             NA     NA -0.3468  0.002
rho           -0.0795  0.029  0.1097  0.014
lambda         0.8236  0.000  0.7959  0.000
```

Finally, continuing with the case of the restricted SUR-GNM model, we can test whether the coefficients of spatial dependence are equal. The **spsur** user can use the function `wald_deltas()` to test the hypothesis,

$$H_0 : \rho_1 = \rho_2 \ \ \text{vs} \ \ H_A : \rho_1 \neq \rho_2$$

So, we would write

```
R> R2 <- matrix(c(1,0,-1,0), nrow = 1)
R> b2 <- matrix(0, ncol = 1)
R> wald_deltas(covid.gnm.rr, R = R2, b = b2)


    Wald test on spatial delta parameters


data:  spain.covid
Wald test = 289.76, df = 1, p-value < 2.2e-16
```

In the case of the hypothesis,

$$H_0 : \lambda_1 = \lambda_2 \ \ \text{vs} \ \ H_A : \lambda_1 \neq \lambda_2$$

the `wald_deltas()` code is similar,

```
R> R2 <- matrix(c(0, 1, 0, -1), nrow = 1)
R> b2 <- matrix(0, ncol = 1)
R> wald_deltas(covid.gnm.rr, R = R2, b = b2)


    Wald test on spatial delta parameters


data:  spain.covid
Wald test = 105.15, df = 1, p-value < 2.2e-16
```

In both cases, the tests confirm that the coefficients of spatial dependence are different, and therefore, the restricted SUR-GNM model is correctly specified.

5.4. Impacts of increased COVID-19 incidence on mobility

To finish the application, it is necessary to interpret the way that changes in the explanatory variables impact both mobility types (inter- and intra-provincial). These impacts can be obtained from the restricted SUR-GNM model. The **spsur** package implements the function `impactspsur()` to compute direct, indirect, and total impacts. The output of this function is a list where each element shows the impacts of one equation,

```
R> impacts.covid.gnm.rr <- impactspsur(covid.gnm.rr, listw = listw, R = 1000)
```

The impacts of inter-provincial mobility are (standard errors, z statistics, and p values, are not included here to save space, they can be obtained by changing `zstats = TRUE`),

```
R> summary(impacts.covid.gnm.rr[[1]], zstats = FALSE, short = TRUE)

Impact measures (mixed, exact):
                        Direct       Indirect        Total
Emergence_1       -0.2860967469   0.0214117578 -0.264684989
EmergenceTotal_1  -0.0849966310   0.0063612302 -0.078635401
Density_1         -0.1047278569   0.0078379343 -0.096889923
Essential_1        0.0005168015   0.0008922296  0.001409031
Incidence_1       -0.0155224499  -0.0175721755 -0.033094625
========================================================
Simulation results ( variance matrix):
```

We will focus the discussion only on the impact of incidence rates on mobility. Similar analysis could be made for the rest of the variables. In the case of the first equation, an increase of one percent in incidence has a direct impact on intra-provincial mobility, with a reduction of 1.55%. This increase in incidence also has an indirect impact (overflow effect), and this effect represents an average reduction of 1.65%. Both direct and indirect impacts involve a total impact of 3.20% on intra-provincial mobility. The summary of the impacts of the second equation for inter-provincial mobility can be obtained with,

```
R> summary(impacts.covid.gnm.rr[[2]], zstats = FALSE, short = TRUE)

Impact measures (mixed, exact):
                        Direct       Indirect        Total
Emergence_2       -0.359389128  -0.043270624 -0.402659752
EmergenceTotal_2  -0.085105845  -0.010246785 -0.095352629
Density_2         -0.074596508  -0.008981455 -0.083577963
Old65_2           -0.347684325  -0.041861360 -0.389545685
Essential_2        0.002301866   0.002936126  0.005237992
Incidence_2       -0.009785371  -0.001178163 -0.010963534
========================================================
Simulation results ( variance matrix):
```

In this case, the impact of a one-percent increase in the COVID-19 incidence rate on intra-provincial mobility is lower than the previous case for inter-provincial mobility. The direct

impact has a reduction on average of 0.98% and the indirect impact, an average of 0.11%. Therefore, the total impact of an increase of one unit in the incidence rate on inter-provincial mobility is 1.09%.

6. Conclusions and directions for future research

This article introduces a new R package, **spsur**, to estimate and make inferences on spatial SUR models. Our intention is to supply researchers interested in spatial econometric methodologies with free, user-friendly software to estimate these types of specifications. **spsur** completes the ecosystem of the R packages, **spdep**, **sphet**, **spse**, **spatialreg**, and **splm**, oriented to deal with spatial econometric specifications.

The package consists of 10 inter-related functions. `lmtestspsur()` tests for the existence of spatial effects in a specified model without spatial mechanisms; the so-called SUR-SIM model. Assuming that some of these tests are statistically significant, three functions, `spsurml()`, `spsur3sls()`, and `spsurtime()` allow the estimation, by ML or 3SLS based on IV, of the preferred spatial specification. The usual methods of extracting information from estimated models, such as `coef()`, `fitted()`, `logLik()`, `plot()`, `print()`, `residuals()`, and `vcov()`, are also included. Four additional functions, `wald_betas()`, `wald_deltas()`, `lr_betas()`, and `anova()` can be used to improve the specification of the model if necessary. The function `impactspsur()` is useful to obtain the spatial multipliers of the variables in the estimated model, and it comes with a collection of measures of statistical significance. Finally, the function `dgp_sur()` is intended to illustrate particular features of spatial SUR specifications as an aid for teaching activities or to implement the results of the Monte Carlo in a more general research project.

Of course, this version of the package should be improved in the future. Included in a list of actions already on the coauthors' table are: (i) the extension of 3SLS to treat endogenous regressors other than the spatial lag of the explained variable and the use of external instruments; (ii) new tests to improve specification; e.g., testing for the instability of the spatial dependence coefficients between equations; (iii) more efficient treatment of the unobserved effects, allowing their estimation, expanding the options to demean the data, (iv) the introduction of time dynamics in a strictly stationary framework; and (v) the consideration of more general SUR models with different types of spatial effects in each equation. Some of this work is already underway and a development version is available from the GitHub repository, including a new function named `spsurgs3sls()` to estimate several spatial SUR models using the GS3SLS. This function is a wrapper of `spreg()` from the **sphet** package, and it implements the GS3SLS estimation method in a SUR framework following the steps explained in Kelejian and Piras (2017, pages 304–305). This function also allows the inclusion of additional endogenous regressors other than the spatial lag of the explained variable.

References

Abdullah M, Dias C, Muley D, Shahin M (2020). "Exploring the Impacts of COVID-19 on Travel Behavior and Mode Preferences." *Transportation Research Interdisciplinary Perspectives*, **8**, 100255. `doi:10.1016/j.trip.2020.100255`.

Agiakloglou C, Tsimpanos A (2021). "Evaluating Information Criteria for Selecting Spatial Processes." *The Annals of Regional Science*, **66**(3), 677–697. `doi:10.1007/s00168-020-01033-y`.

Anselin L (1988a). *Spatial Econometrics: Methods and Models*. Studies in Operational Regional Science. Springer-Verlag, Dordrecht. `doi:10.1007/978-94-015-7799-1`.

Anselin L (1988b). "A Test for Spatial Autocorrelation in Seemingly Unrelated Regressions." *Economics Letters*, **28**(4), 335–341. `doi:10.1016/0165-1765(88)90009-2`.

Anselin L (1992). **SpaceStat** *Tutorial. A Workbook for Using* **SpaceStat** *in the Analysis of Spatial Data*. Regional Research Institute, West Virginia University, Morgantown. URL `https://web.worldbank.org/archive/website01407/WEB/IMAGES/ANSELIN_.PDF`.

Anselin L (2016). "Estimation and Testing in the Spatial Seemingly Unrelated Regression (SUR) Model." *Technical report*, Geoda Center for Geospatial Analysis and Computation, Arizona State University. `doi:10.13140/rg.2.2.15925.40163`. Working Paper 2016-01.

Anselin L, Bera AK, Florax R, Yoon MJ (1996). "Simple Diagnostic Tests for Spatial Dependence." *Regional Science and Urban Economics*, **26**(1), 77–104. `doi:10.1016/0166-0462(95)02111-6`.

Anselin L, Le Gallo J, Jayet H (2008). "Spatial Panel Econometrics." In L Matyas, P Sevestre (eds.), *The Econometrics of Panel Data. Fundamentals and Recent Developments in Theory and Practice*, chapter 19, pp. 625–660. Springer-Verlag. `doi:10.1007/978-3-540-75892-1_19`.

Baller RD, Anselin L, Messner SF, Deane G, Hawkins DF (2001). "Structural Covariates of US County Homicide Rates: Incorporating Spatial Effects." *Criminology*, **39**(3), 561–588. `doi:10.1111/j.1745-9125.2001.tb00933.x`.

Baltagi BH, Bresson G (2011). "Maximum Likelihood Estimation and Lagrange Multiplier Tests for Panel Seemingly Unrelated Regressions with Spatial Lag and Spatial Errors: An Application to Hedonic Housing Prices in Paris." *Journal of Urban Economics*, **69**(1), 24–42. `doi:10.1016/j.jue.2010.08.007`.

Baltagi BH, Pirotte A (2011). "Seemingly Unrelated Regressions with Spatial Error Components." *Empirical Economics*, **40**(1), 5–49. `doi:10.1007/s00181-010-0373-8`.

Bera AK, Yoon MJ (1993). "Specification Testing with Locally Misspecified Alternatives." *Econometric Theory*, **9**(4), 649–658. `doi:10.1017/s0266466600008021`.

Bivand R, Hauke J, Kossowski T (2013a). "Computing the Jacobian in Gaussian Spatial Autoregressive Models: An Illustrated Comparison of Available Methods." *Geographical Analysis*, **45**(2), 150–179. `doi:10.1111/gean.12008`.

Bivand R, Millo G, Piras G (2021). "A Review of Software for Spatial Econometrics in R." *Mathematics*, **9**(11), 1276. `doi:10.3390/math9111276`.

Bivand R, Pebesma E, Gómez-Rubio V (2013b). *Applied Spatial Data Analysis with R.* 2nd edition. Springer-Verlag. URL `https://asdar-book.org/`.

Bivand R, Piras G (2015). "Comparing Implementations of Estimation Methods for Spatial Econometrics." *Journal of Statistical Software*, **63**(18), 1–36. `doi:10.18637/jss.v063.i18`.

Bivand R, Wong DWS (2018). "Comparing Implementations of Global and Local Indicators of Spatial Association." *TEST*, **27**(3), 716–748. `doi:10.1007/s11749-018-0599-x`.

Breusch TS, Pagan AR (1980). "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics." *The Review of Economic Studies*, **47**(1), 239–253. `doi:10.2307/2297111`.

Brodeur A, Grigoryeva I, Kattan L (2021). "Stay-at-Home Orders, Social Distancing, and Trust." *Journal of Population Economics*, **34**(4), 1321–1354. `doi:10.1007/s00148-021-00848-z`.

Burridge P, Elhorst JP, Zigova K (2016). "Group Interaction in Research and the Use of General Nesting Spatial Models." In *Spatial Econometrics: Qualitative and Limited Dependent Variables*, volume 37, pp. 223–258. Emerald Group Publishing Limited. `doi:10.1108/s0731-905320160000037016`.

Cotteleer G, Van Kooten GC (2012). "Expert Opinion versus Actual Transaction Evidence in the Valuation of Non-Market Amenities." *Economic Modelling*, **29**(1), 32–40. `doi:10.1016/j.econmod.2011.02.041`.

Davidson R, MacKinnon JG (1993). *Estimation and Inference in Econometrics.* Oxford University Press. URL `http://qed.econ.queensu.ca/pub/dm-book/`.

Egger P, Pfaffermayr M (2004). "Distance, Trade and FDI: A Hausman–Taylor SUR Approach." *Journal of Applied Econometrics*, **19**(2), 227–246. `doi:10.1002/jae.721`.

Elhorst JP (2014). *Spatial Econometrics. From Cross-Sectional to Spatial Panels.* Springer-Verlag. URL `https://link.springer.com/book/10.1007/978-3-642-40340-8`.

Engle S, Stromme J, Zhou A (2020). "Staying at Home: Mobility Effects of COVID-19." *SSRN Electronic Journal.* `doi:10.2139/ssrn.3565703`.

Fingleton B (2007). "A Multi-Equation Spatial Econometric Model, with Application to EU Manufacturing Productivity Growth." *Journal of Geographical Systems*, **9**(2), 119–144. `doi:10.1007/s10109-006-0038-5`.

Franklin J, Rey SJ (2016). "Heterogeneous Tree Recruitment Following Disturbance in Insular Tropical Forest, Kingdom of Tonga." *Journal of Tropical Ecology*, **32**(6), 536. `doi:10.1017/s0266467416000456`.

Harris R, Moffit J, Kravtsova V (2011). "In Search of W." *Spatial Economic Analysis*, **6**, 249–270. `doi:10.1080/17421772.2011.586721`.

He L, Páez A, Liu D, Jiang S (2015). "Temporal Stability of Model Parameters in Crime Rate Analysis: An Empirical Examination." *Applied Geography*, **58**, 141–152. `doi:10.1016/j.apgeog.2015.02.002`.

Izón GM, Hand MS, Mccollum DW, Thacher JA, Berrens RP (2016). "Proximity to Natural Amenities: a Seemingly Unrelated Hedonic Regression Model with Spatial Durbin and Spatial Error Processes." *Growth and Change*, **47**(4), 461–480. `doi:10.1111/grow.12147`.

Kakamu K, Polasek W, Wago H (2012). "Production Technology and Agglomeration for Japanese Prefectures during 1991–2000." *Papers in Regional Science*, **91**(1), 29–41. `doi:10.1111/j.1435-5957.2011.00360.x`.

Kapoor M, Kelejian HH, Prucha IR (2007). "Panel Data Models with Spatially Correlated Error Components." *Journal of Econometrics*, **140**(1), 97–130. `doi:10.1016/j.jeconom.2006.09.004`.

Kelejian H, Piras G (2017). *Spatial Econometrics*. Academic Press. `doi:10.1016/C2016-0-04332-2`.

Kelejian H, Piras G (2020). "Spillover Effects in Spatial Models: Generalizations and Extensions." *Journal of Regional Science*, **60**(3), 425–442. `doi:10.1111/jors.12476`.

Kelejian HH, Prucha IR (1999). "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review*, **40**(2), 509–533. `doi:10.1111/1468-2354.00027`.

Kelejian HH, Prucha IR (2004). "Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations." *Journal of Econometrics*, **118**(1-2), 27–50. `doi:10.1016/s0304-4076(03)00133-7`.

Lauridsen J, Bech M, López FA, Maté M (2010). "A Spatiotemporal Analysis of Public Pharmaceutical Expenditure." *The Annals of Regional Science*, **44**(2), 299–314. `doi:10.1007/s00168-008-0259-2`.

Le Gallo J, Chasco C (2008). "Spatial Analysis of Urban Growth in Spain, 1900–2001." *Empirical Economics*, **34**(1), 59–80. `doi:10.1007/s00181-007-0150-5`.

Le Gallo J, Dall'erba S (2006). "Evaluating the Temporal and Spatial Heterogeneity of the European Convergence Process, 1980–1999." *Journal of Regional Science*, **46**(2), 269–288. `doi:10.1111/j.0022-4146.2006.00441.x`.

Lee LF (2004). "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models." *Econometrica*, **72**(6), 1899–1925. `doi:10.1111/j.1468-0262.2004.00558.x`.

LeSage J, Pace RK (2009). *Introduction to Spatial Econometrics*. Chapman & Hall/CRC. `doi:10.1201/9781420064254`.

López FA, Martínez-Ortiz PJ, Cegarra-Navarro JG (2017). "Spatial Spillovers in Public Expenditure on a Municipal Level in Spain." *The Annals of Regional Science*, **58**(1), 39–65. `doi:10.1007/s00168-016-0780-7`.

López FA, Mínguez R, Mur J (2020). "ML versus IV Estimates of Spatial SUR Models: Evidence from the Case of Airbnb in Madrid Urban Area." *The Annals of Regional Science*, **64**(2), 313–347. `doi:10.1007/s00168-019-00914-1`.

López FA, Mínguez R, Mur J (2022). **spsur***: Spatial Seemingly Unrelated Regression Models*. R package version 1.0.2.5, URL `https://CRAN.R-project.org/package=spsur`.

López FA, Mur J, Angulo A (2014). "Spatial Model Selection Strategies in a SUR Framework. The Case of Regional Productivity in EU." *The Annals of Regional Science*, **53**(1), 197–220. `doi:10.1007/s00168-014-0624-2`.

Lundberg J (2006). "Spatial Interaction Model of Spillovers from Locally Provided Public Services." *Regional Studies*, **40**(6), 631–644. `doi:10.1080/00343400600868788`.

Millo G, Piras G (2012). "**splm**: Spatial Panel Data Models in R." *Journal of Statistical Software*, **47**(1), 1–38. `doi:10.18637/jss.v047.i01`.

Moscone F, Tosetti E, Knapp M (2007). "SUR Model with Spatial Effects: An Application to Mental Health Expenditure." *Health Economics*, **16**(12), 1403–1408. `doi:10.1002/hec.1229`.

Mur J, Angulo A (2009). "Model Selection Strategies in a Spatial Setting: Some Additional Results." *Regional Science and Urban Economics*, **39**(2), 200–213. `doi:10.1016/j.regsciurbeco.2008.05.018`.

Mur J, López FA, Herrera M (2010). "Testing for Spatial Effects in Seemingly Unrelated Regressions." *Spatial Economic Analysis*, **5**(4), 399–440. `doi:10.1080/17421772.2010.516443`.

Páez A, López FA, Menezes T, Cavalcanti R, Pitta MGDR (2021). "A Spatio-Temporal Analysis of the Environmental Correlates of COVID-19 Incidence in Spain." *Geographical Analysis*, **53**(3), 397–421. `doi:10.1111/gean.12241`.

Pebesma E (2018). "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal*, **10**(1), 439–446. `doi:10.32614/rj-2018-009`.

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

Piras G (2018). **spse***: Spatial Simultaneous Equations and Spatial SUR Model*. R package version 0.0-2, URL `https://R-forge.R-project.org/projects/spse/`.

Piras G, Prucha IR (2014). "On the Finite Sample Properties of Pre-Test Estimators of Spatial Models." *Regional Science and Urban Economics*, **46**, 103–115. `doi:10.1016/j.regsciurbeco.2014.03.002`.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Rey SJ, Anselin L (2009). "**PySAL**: A Python Library of Spatial Analytical Methods." In *Handbook of Applied Spatial Analysis*, pp. 175–193. Springer-Verlag. `doi:10.1007/978-3-642-03647-7_11`.

Rey SJ, Montouri BD (1999). "US Regional Income Convergence: A Spatial Econometric Perspective." *Regional Studies*, **33**(2), 143–156. `doi:10.1080/00343409950122945`.

Tupy IS, Silva FF, Amaral PVM, Cavalcante ATM (2021). "The Spatial Features of Recent Crises in a Developing Country: Analysing Regional Economic Resilience for the Brazilian Case." *Regional Studies*, **55**(4), 693–706. `doi:10.1080/00343404.2020.1851025`.

Vega SH, Elhorst JP (2015). "The SLX Model." *Journal of Regional Science*, **55**(3), 339–363. `doi:10.1111/jors.12188`.

Wang X, Kockelman KM (2007). "Specification and Estimation of a Spatially and Temporally Autocorrelated Seemingly Unrelated Regression Model: Application to Crash Rates in China." *Transportation*, **34**(3), 281–300. `doi:10.1007/s11116-007-9117-9`.

Wooldridge JM (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd edition. Massachusets Institut of Technology Press, Cambridge, Massachussetts. URL `https://mitpress.mit.edu/9780262232586/`.

Zeileis A, Croissant Y (2010). "Extended Model Formulas in R: Multiple Parts and Multiple Responses." *Journal of Statistical Software*, **34**(1), 1–13. `doi:10.18637/jss.v034.i01`.

Zellner A (1962). "An Efficient Method of Estimating Seemingly Unrelated Regressions and Test of Aggregation Bias." *Journal of the American Statistical Association*, **57**, 500–509. `doi:10.2307/2281644`.

Affiliation:

Román Mínguez
University of Castilla-La Mancha
Avenida de Los Alfares, 42. 16002 Cuenca, Spain
E-mail: `roman.minguez@uclm.es`
URL: `https://github.com/rominsal`

Fernando A. López
Technical University of Cartagena
Facultad de CC de la Empresa
C/ Real, 3. 30201 Cartagena, Murcia, Spain
E-mail: `fernando.lopez@upct.es`

Jesús Mur
University of Zaragoza
Department of Economic Analysis
C/ María de Luna s/n 50018 Zaragoza, Spain
E-mail: jmur@unizar.es