

Spatial SUR models: Specification, Testing and Selection.

Jesús Mur (*), *Fernando López* (**)

May 2010

(*) Department of Economic Analysis
University of Zaragoza
Gran Vía, 2-4. (50005). Zaragoza. Spain.
e-mail: jmur@unizar.es

(**) Department of Quantitative Methods and Computing
Technical University of Cartagena.
Calle Real, 3. 30201. Cartagena. Spain.
e-mail: fernando.lopez@upct.es

Abstract

Our paper focuses on the case of SUR models with spatial effects. Specifically, the problem that we pose is testing for the presence of spatial effects in these multivariate models, considering that one of our main interests is to deal with the question of selecting the most adequate specification for the data. In order to do this, we obtain several tests that check for the fundamental hypothesis of the model, including the assumption of time stability of the spatial dependence mechanisms. We solve the discussion in a maximum-likelihood framework because this facilitates the obtaining of appropriate tests. The second part of the paper contains a Monte Carlo experiment in order to study the behaviour of the two most popular model selection strategies, Specific-to-General or General-to-Specific, using the collection of tests that have been proposed.

Keywords: Spatial dependence; Seemingly Unrelated Regressions; Model Selection; Monte Carlo.

JEL Classification: C21; C50; R15

Acknowledgements: This work has been carried out with the financial support of project ECO2009-10534/ECON of the Ministerio de Ciencia y Tecnología of the Reino de España.

1. Introduction

The Seemingly Unrelated Regression equations model (SUR from now on) is a popular multivariate econometric formulation employed in very different fields, included the analysis of spatial data. The basis of this approach are well-known since the initial works of Zellner (1962), Theil (1971), Malinvaud (1970), Schmidt (1976) or Dwivedi and Srivastava (1978). Almost every textbook in Econometrics includes a discussion about SUR, which it is available in the most popular econometric computer programmes. It hardly requires any further justification.

Anselin (1988a) introduces the term of spatial SUR in reference to a special case of a more general space-time model. According to him, the spatial SUR model '*consists of an equation for each time period which is estimated for a cross-section of spatial units*' (p. 141). A characteristic of this approach is the existence of a limited heterogeneity. In fact, the regression coefficients are assumed to be the same across individuals whereas individual unobserved effects are excluded. Different spatial mechanisms may intervene in each equation (in example, intra-equation spatial error autocorrelation, spatially autoregressive processes, etc). Usually, the cross-sectional dimension of the sample (say R , number of individuals) is greater than the temporal dimension (say T , number of cross sections). Rey and Montouri (1999), Fingleton (2001, 2007), Egger and Pfaffermayr (2004), Moscone et al. (2007), LeGallo and Chasco (2008) or Lauridsen et al (2010) employ a similar framework. If we need a more flexible model, in the sense of a greater heterogeneity, the next step should consist of a panel data model, with fixed or random unobservable effects (Anselin et al, 2007, Kapoor et al, 2007, Baltagi, 2008) and certain spatial structure (Elhorst, 2003, 2005, 2008).

A different situation is produced when the time dimension is greater than the cross-sectional dimension (the ratio R/T goes to zero). In this case the specification rests, mainly, in the time dimension of the model. Heterogeneity emerges as an important question given that there are few individuals, being possible to develop an equation for each individual in addition to the usual interaction mechanisms. Arora and Brown (1977), Hordijk and Nijkamp (1977), Hordijk (1979) or White and Hewings (1982) follow this approach. In this setting (rich temporal information but few spatial details), the problem of specifying a spatial weighting matrix is of minor importance because several consistent estimator can be obtained from the data. This is the way followed by Conley (1999), Chen and Conley (2001), Coakley et al (2002), Pesaran (2005) or Conley and Molinari (2008), among others. Problems appear when R increases at a rate similar to T . As indicated by Driscoll and Kraay (1998), in this case it is necessary to introduce restrictions on the number of parameters to keep under control the dimensions of the problem. Carlino and deFina (1999), Di Giacinto (2003, 2006), Badinger et al (2004), or Beenstock and Felsenstein (2007, 2008) present

different application in this line that may be called of Spatial Vector Autoregressive models, SpVAR.

In this paper we address the case of a SUR model, which involves spatial data, with spatial effects, under the configuration of a finite T , a large R and a limited heterogeneity among individuals. As said, our main problem is testing for the presence of spatial effects in the specification and, then, to identify the type of spatial process that seems to be more adequate for the data.

We are going to use a maximum-likelihood approach which facilitates the obtaining of simple tests, generally well-behaved in a small sample context, based on the principle of the Lagrange Multiplier. The work of Kelejian and Prucha (2004), dealing with systems of simultaneous equations, including different spatial mechanisms, is very close to our own work. These authors obtain a limited and a full information estimator, based on an approximation to the optimal set of instruments for different cases of interest. As a complement of their work, Kelejian and Prucha (2004, p.40) demand '*the development of further tests of hypotheses in a spatial system framework*'. Baltagi and Pirotte (2009) focus on the estimation of properly SUR models with spatial error components, examining both maximum likelihood and generalized moment methods. According to their simulation experiment, the behaviour of both estimation algorithms is similar, conditioned to the correct specification of the model. This means that some specification tests are needed to guide the estimation. In sum, as expressed by LeGallo and Dall'erna (2006, p. 279), the current situation is not fully satisfactory: '*For our SUR specification with spatial autocorrelation and spatial regimes, no specification procedure has been formally suggested*'.

The paper contains seven sections. In the second section we specify a general SUR model with spatial effects, that we call SUR-SARAR model. In the third section we develop a maximum-likelihood framework in order to test for the presence of spatial effects in this specification. The fourth section introduces, in a SUR context, the well-known robust and marginal Lagrange Multipliers. Some extensions follow in the fifth section. In the sixth we solve a Monte Carlo experiment directly aimed at study the behaviour of the two most popular model selection strategies, Specific-to-General and General-to-Specific, using the collection of tests developed in the previous sections. Finally, the seventh section comments the main conclusion reached in our work.

2. Specification of the model.

In the rest of the paper, we are going to deal with a SUR model which includes some spatial mechanisms, like the following:

$$\begin{aligned}
& \left. \begin{aligned} y_{gt} &= \lambda_g \mathbf{W}_1 y_{gt} + x_{gt} \beta_g + u_{gt} & \Rightarrow \mathbf{A}_g y_{gt} &= x_{gt} \beta_g + u_{gt} \\ u_{gt} &= \rho_g \mathbf{W}_2 u_{gt} + \varepsilon_{gt} & \Rightarrow \mathbf{B}_g u_{gt} &= \varepsilon_{gt} \end{aligned} \right\} \\
& \qquad \qquad \qquad g = 1, 2, \dots, G \\
& E[\varepsilon_{gt}] = 0 \quad E[\varepsilon_{gt} \varepsilon'_{ht}] = \sigma_{gh} \mathbf{I}_R \\
& \mathbf{A}_{gt} = \mathbf{I}_R - \lambda_g \mathbf{W}_1 \quad \mathbf{B}_{gt} = \mathbf{I}_R - \rho_g \mathbf{W}_2
\end{aligned} \tag{1}$$

y_{gt} , u_{gt} and ε_{gt} are $(R \times 1)$ vectors, x_{gt} is a matrix of exogenous variables of order $(R \times k_g)$, β_g is a vector of parameters of order $(k_g \times 1)$, λ_g and ρ_g are two scalars¹, \mathbf{I}_R is the identity matrix of order $(R \times R)$ and \mathbf{W}_1 and \mathbf{W}_2 are two known weighting matrices of order $(R \times R)$. In terms of Kelejian and Prucha (2001), this is a SUR model with Spatial ARAR(1,1) process² (SUR-SARAR for the sake of brevity).

The observations are dated in period t ($t=1, 2, \dots, T$) and they proceed from R individuals (spatial units), which are spatially distributed. The distinctive feature of model (1) is that there exist spatial spillover effects. The weighting matrices \mathbf{W}_1 and \mathbf{W}_2 describe how these spillovers are produced. The specification of these matrices is rather arbitrary (Haining, 2001). The elements of the main diagonal are zero whereas a nonzero value outside this diagonal reflects that the two observations are (geographically, sociologically, technologically, etc) neighbours. For simplicity's sake, we will assume that the two weighting matrices coincide ($\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{W}$).

The spatial SUR model of Anselin (1988a) corresponds to the case of $G=1$, as discussed in Mur and López (2009). This model is, in fact, a panel data model given that it contains R individuals, T cross-sections and 1 equation. The case of R individuals, G equations and 1 cross-section is more in line with the SUR spirit (Elhorst, 2003); obviously, the two approaches can be combined. In what follows, we maintain the general framework of (1). In matrix terms:

$$\left. \begin{aligned} \mathbf{A} \mathbf{y} &= \mathbf{X} \beta + \mathbf{u} \\ \mathbf{B} \mathbf{u} &= \varepsilon \\ \varepsilon &\sim N(0, \Omega) \end{aligned} \right\} \tag{2}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_T \end{bmatrix}; \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \dots \\ \mathbf{X}_T \end{bmatrix}; \quad \mathbf{X}_t = \begin{bmatrix} x_{1t} & 0 & \dots & 0 \\ 0 & x_{2t} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & x_{Gt} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_G \end{bmatrix}; \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_T \end{bmatrix}; \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_T \end{bmatrix}$$

TGRx1 TGRxk GRxk kx1 TGRx1 TGRx1

Where $k = \sum_{g=1}^G k_g$; $\mathbf{A} = \mathbf{I}_T \otimes [\mathbf{I}_{GR} - \Lambda \otimes \mathbf{W}]$ and $\mathbf{B} = \mathbf{I}_T \otimes [\mathbf{I}_{GR} - \Upsilon \otimes \mathbf{W}]$, Λ (respectively Υ) is an $(G \times G)$ diagonal matrix with the parameters λ_g (respectively ρ_g) and \otimes is the Kronecker

¹ The subindex t used in the specification of matrices \mathbf{A}_{gt} and \mathbf{B}_{gt} means that we do not exclude the possibility of having time varying spatial dependence parameters, λ_{gt} and/or ρ_{gt} .

² That is, with an autoregressive structure both in the main equation and in the equation of the errors. Spatial moving average structures are also of interest but they are not included by length restrictions.

product. Moreover $\Omega = I_T \otimes \Sigma_\varepsilon \otimes I_R$ where $\Sigma_\varepsilon = [\sigma_{\varepsilon ij}; i, j = 1, 2, \dots, G]$ is a $G \times G$ matrix. We assume normality in the error terms.

It is important to consider the followings remarks:

- We order the sampling information, first, temporarily; then, we sort each cross section by equation and, finally, by individuals.
- Different set of regressors may intervene in each equation although, in order to simplify, we assume that the vector of parameters of each equation ($\beta_g, g=1, 2, \dots, G$) is the same across cross-sections and individuals.
- Accordingly, we assume that the parameters of spatial dependence ($\lambda_g, \rho_g, g=1, 2, \dots, G$) are also constant in time but that may vary between different equations.
- The SUR effect is due to the fact that the same individual (the spatial unit) decides, simultaneously, about G different problems (equations). This situation is similar to the existence of an unobserved random effect which affects in the same manner to all individuals in each equation; the random effect varies between equations. Baltagi and Pirotte (2009) go a step further by introducing individual unobserved random effects, which allow them for a greater heterogeneity:

$$\left. \begin{aligned} \mathbf{A}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{B}\mathbf{u} &= \boldsymbol{\varepsilon} + \boldsymbol{\mu} \end{aligned} \right\}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_T \end{bmatrix}_{TGR \times 1}; \boldsymbol{\mu} = [I_T \otimes I_{RG}] \boldsymbol{\eta}; \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \dots \\ \eta_G \end{bmatrix}_{GR \times 1}; \boldsymbol{\eta}_g = \begin{bmatrix} \eta_{1g} \\ \eta_{2g} \\ \dots \\ \eta_{Rg} \end{bmatrix}_{R \times 1} \quad (3)$$

$$\begin{aligned} E[\boldsymbol{\varepsilon}] &= 0 & E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] &= \Omega = \Sigma_\varepsilon \otimes I_{TR} \\ E[\boldsymbol{\mu}] &= 0 & E[\boldsymbol{\mu}\boldsymbol{\mu}'] &= \Xi = J_T \otimes [\Sigma_\eta \otimes I_R] \\ E[\boldsymbol{\varepsilon}\boldsymbol{\mu}'] &= 0 \end{aligned}$$

η_{rg} is the effect associated to the r -th individual in the g -th equation, $\boldsymbol{\eta}'_r = (\eta_{1r}; \eta_{2r}; \dots; \eta_{Gr})$ is the vector of effects for the r -th individual, $\Sigma_\eta = E[\boldsymbol{\eta}_r \boldsymbol{\eta}'_r] = [\sigma_{\eta ij}; i, j = 1, 2, \dots, G]; \forall r$ a $G \times G$ covariance matrix and $J_T = I_T I_T'$ a $T \times T$ matrix of ones. Both terms are assumed to be orthogonal³.

Following Kelejian and Prucha (2004), in the specification of (2) we assume the usual set of hypothesis: a) The spatial lag is well-behaved in the sense that the elements in the main diagonal of

³ Baltagi and Pirotte (2009) order the information, in first place, by equation, then by time and finally by individuals. Moreover, they do not use autoregressive mechanisms in the main equation; only autoregressive or moving average processes in the equation of the errors.

the weighting matrices \mathbf{W}_1 and \mathbf{W}_2 are zero, the matrices \mathbf{A}_{gt} and \mathbf{B}_{gt} are nonsingular and the row and column sums of the matrices \mathbf{W}_1 , \mathbf{W}_2 , \mathbf{A}_{gt}^{-1} and \mathbf{B}_{gt}^{-1} are uniformly bounded in absolute value; b) The matrix of exogenous variables, \mathbf{X} , is well-behaved given that it is of full column rank and its elements are uniformly bounded in absolute value; c) The innovation terms $\varepsilon_{rt} = [\varepsilon_{1rt}, \varepsilon_{2rt}, \dots, \varepsilon_{Grt}]'$ for each r and t , are distributed independently and identically with zero mean and a nonsingular variance covariance matrix Σ_ε . We add the hypothesis of normality to facilitate the maximum-likelihood approach.

Next we are going to discuss the case of the SUR-SARAR model as presented in (2). Moreover, we present the cases of the SUR-SLM model (obtained after introducing the restrictions $\rho_g=0, \forall g$ in (2)) and the SUR-SEM model (the restrictions in (2) are $\lambda_g=0, \forall g$).

3. Testing for the presence of spatial effects in a SUR-SARAR model.

The logarithm of the likelihood function of the SUR-SARAR of (2) is the following:

$$l(y; \theta) = -\frac{RTG}{2} \ln(2\pi) - \frac{RT}{2} \ln|\Sigma| + T \left[\sum_{g=1}^G \ln|\mathbf{B}_g| + \sum_{g=1}^G \ln|\mathbf{A}_g| \right] - \frac{(\mathbf{A}y - \mathbf{X}\beta)' \mathbf{B}' \Omega^{-1} \mathbf{B} (\mathbf{A}y - \mathbf{X}\beta)}{2} \quad (4)$$

where $\theta' = [\beta'; \lambda_1; \dots; \lambda_G; \rho_1; \dots; \rho_G; \sigma_{ij}]$ is the vector of parameters, of order $(k+2G+G(G+1)/2) \times 1$ and $\Omega^{-1} = \mathbf{I}_T \otimes \Sigma^{-1} \otimes \mathbf{I}_R$. The ML estimation results in a nonlinear optimization problem which can be solved applying standard numerical techniques (Wang and Kockelman 2007). In the Appendix, Section A.I we include more details.

The null hypothesis of absence of spatial effects in the SUR model of (2) is:

$$H_0: \lambda_g = \rho_g = 0 (\forall g) \quad \text{vs} \quad H_A: \text{No } H_0 \quad (5)$$

After a few calculi the final expression of the corresponding Lagrange Multipliers test⁴ is:

$$\mathbf{LM}_{\text{SARAR}}^{\text{SUR}} = \begin{bmatrix} \mathbf{g}'(\lambda)_{|H_0} & \mathbf{g}'(\rho)_{|H_0} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{\lambda\lambda} - \mathbf{I}_{\lambda\beta} \mathbf{I}_{\beta\beta}^{-1} \mathbf{I}_{\beta\lambda} & \mathbf{I}_{\lambda\rho} \\ \mathbf{I}_{\rho\lambda} & \mathbf{I}_{\rho\rho} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}(\lambda)_{|H_0} \\ \mathbf{g}(\rho)_{|H_0} \end{bmatrix} \underset{\text{as}}{\sim} \chi^2(2G) \quad (6)$$

⁴ In what follows we will use a compact standard notation:

$$\mathbf{LM} = \begin{bmatrix} \mathbf{g}(\theta)_{|H_0} \end{bmatrix}' \begin{bmatrix} \mathbf{I}(\theta)_{|H_0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}(\theta)_{|H_0} \end{bmatrix} \underset{\text{as}}{\sim} \chi^2(df)$$

where $\mathbf{g}(\theta)$ is the score (vector of first derivatives of the likelihood function), $\mathbf{I}(\theta)$ the information matrix, df means degrees of freedom and ' $|H_0$ ' means evaluated under the hypothesis H_0 .

with $\mathbf{g}_{(\lambda)}|_{H_0} = \hat{\mathbf{u}}' \left[\mathbf{I}_T \otimes \left(\sum_{g=1, \dots, G}^{-1} \mathbf{E}^{gg} \right) \otimes \mathbf{W} \right] \mathbf{y}$ and $\mathbf{g}_{(\rho)}|_{H_0} = \hat{\mathbf{u}}' \left[\mathbf{I}_T \otimes \left(\sum_{g=1, \dots, G}^{-1} \mathbf{E}^{gg} \right) \otimes \mathbf{W} \right] \hat{\mathbf{u}}$ where $\hat{\mathbf{u}}$ is the

(TGRx1) vector of residuals of the SUR model, estimated in the absence of spatial effects, and \mathbf{E}^{gg} a (GxG) matrix whose elements are all zero except the (g,g) which is 1.

The SUR-SLM model is a particular case of the SUR-SARAR of (2), which includes lags of the endogenous variable on the right hand side of the equation but there is no spatial structure in the equation of the errors. That is:

$$\left. \begin{aligned} \mathbf{A}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathbf{N}(0, \boldsymbol{\Omega}) \end{aligned} \right\} \quad (7)$$

where $\mathbf{A} = \mathbf{I}_T \otimes [\mathbf{I}_{GR} - \Lambda \otimes \mathbf{W}]$ and $\boldsymbol{\Omega}$ is the same matrix of (2). The hypothesis of no spatial effects becomes:

$$H_0 : \lambda_g = 0 \quad (\forall g) \quad \text{vs} \quad H_A : \text{No } H_0 \quad (8)$$

The Lagrange Multiplier is (Section A.II of the Appendix for the details):

$$\mathbf{LM}_{\text{SLM}}^{\text{SUR}} = \mathbf{g}'_{(\lambda)}|_{H_0} \left[\mathbf{I}_{\lambda\lambda} - \mathbf{I}_{\lambda\beta} \mathbf{I}_{\beta\beta}^{-1} \mathbf{I}_{\beta\lambda} \right]^{-1} \mathbf{g}_{(\lambda)}|_{H_0} \sim \chi^2(G) \quad (9)$$

Finally, the SUR-SEM model includes spatial dependence in the equation of the errors but there are no spatial lags of the endogenous on the right hand side of the main equation:

$$\left. \begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{B}\mathbf{u} &= \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathbf{N}(0, \boldsymbol{\Omega}) \end{aligned} \right\} \quad (10)$$

This model is well known since the seminal work of Anselin (1988b). We include it here only to give a more complete view of the discussion. The null hypothesis of no spatial effects is:

$$H_0 : \rho_g = 0 \quad (\forall g) \quad \text{vs} \quad H_A : \text{No } H_0 \quad (11)$$

The obtaining of the corresponding Multiplier is straightforward (Section A.III of the Appendix):

$$\mathbf{LM}_{\text{SEM}}^{\text{SUR}} = \mathbf{g}'_{(\rho)}|_{H_0} \left[\mathbf{I}_{\rho\rho} \right]^{-1} \mathbf{g}_{(\rho)}|_{H_0} \sim \chi^2(G) \quad (12)$$

4. The robust and the marginal Lagrange Multipliers: SUR-SLM vs SUR-SEM models.

The tests of the last section shall help us to improve the specification of a spatial SUR model although, probably, they will not be definite. The problem with these three tests is that they are not robust to local misspecification errors in the alternative (Davidson and McKinnon, 1996). Consequently, they are not good instruments in order to identify the type of model that has intervened in the data generating process (DGP from now on). This lack of robustness confer great importance to the work of Bera and Yoon (1993), which obtain the correction needed for the raw

Multipliers in order to behave properly. Next, we present the case of \mathbf{LM}_{SLM}^{SUR} and \mathbf{LM}_{SEM}^{SUR} tests, following Anselin et al (1996).

The likelihood function of the SUR-SARAR model, $L[\varphi; \lambda; \rho]$, depends on three groups of parameters: those associated to the basic SUR structure, $\varphi' = [\beta'; \sigma_{ij}]$, those related to the spatial lag of the endogenous variable, $\lambda' = [\lambda_1; \dots; \lambda_G]$, and those that introduce spatial dependence into the error terms, $\rho' = [\rho_1; \dots; \rho_G]$. The \mathbf{LM}_{SLM}^{SUR} tests the hypothesis that vector λ is zero assuming implicitly that vector ρ is zero. The null hypothesis of the \mathbf{LM}_{SEM}^{SUR} refers to vector ρ , assuming that vector λ is zero. As its well documented in the literature (see, for example, Anselin et al, 1996, Florax et al, 2003, or Mur and Angulo, 2008), both tests will behave properly only if the DGP has been correctly specified

Let us assume that the DGP is $L_1[\varphi; \lambda]$, then $\mathbf{LM}_{SLM}^{SUR} \xrightarrow{D} \chi^2(G; \alpha)$ where G refers to the degrees of freedom and α is a noncentrality parameter. If the null hypothesis, $H_0: \lambda=0$, is true then the noncentrality parameter will be zero. On the other hand, if the DGP is of the type $H_A: \lambda = \xi_\lambda / \sqrt{R}$ with $\xi_\lambda \neq 0$ and finite, the noncentrality parameter becomes $\alpha = \xi_\lambda' I_{\lambda \cdot \varphi} \xi_\lambda$ with $I_{\lambda \cdot \varphi} = I_{\lambda \lambda} - I_{\lambda \varphi} I_{\varphi \varphi}^{-1} I_{\varphi \lambda}$. It is simple to verify (Anselin et al, 1996) that $\alpha > 0$, which increases the power of the \mathbf{LM}_{SLM}^{SUR} test.

Problems appear when the DGP is $L_2[\varphi; \rho]$ where, effectively, λ is zero. If we, erroneously, assume that the DGP is $L_1[\varphi; \lambda]$, the impact of vector ρ will be omitted which is a source of confusion. Let us assume that $\rho = \xi_\rho / \sqrt{R}$ and $\xi_\rho \neq 0$; the null hypothesis, $H_0: \lambda = 0$, is true and, as expected, $\mathbf{LM}_{SLM}^{SUR} \xrightarrow{D} \chi^2(G; \alpha)$. However, the noncentrality parameter will not be zero but $\alpha = \xi_\rho' I_{\lambda \cdot \varphi} I_{\lambda \cdot \varphi}^{-1} I_{\rho \lambda \cdot \varphi} \xi_\rho$ with $I_{\lambda \cdot \varphi} = I_{\lambda \rho} - I_{\lambda \varphi} I_{\varphi \varphi}^{-1} I_{\varphi \rho}$. In sum, if the DGP is of the SEM type, the \mathbf{LM}_{SLM}^{SUR} test will tend to reject, unduly, the null hypothesis. By a similar reasoning, it can be shown that something similar happens with the \mathbf{LM}_{SEM}^{SUR} test: it will tend to reject, unduly, the null hypothesis ($H_0: \rho = 0$) if the DGP is of the SLM type.

Bera and Yoon (1993) propose the development of ‘*size-resistant*’ tests in order to assure the desired Type I error level even under local misspecifications of the alternative hypothesis. A natural solution is to adjust both the score and the information matrix that appear in the raw Lagrange Multipliers. The objective is to ‘*robustify*’ these Multipliers and the procedure is

relatively simple (here we are using the restricted version of the model, $H_0 : \lambda = \rho = 0$). As shown by Anselin et al. (1996), this adjustment works well in a spatial context. In our case and using the results of Section A.I in the Appendix:

$$\left. \begin{array}{l} H_0 : \lambda_g = 0; \forall g \\ H_A : \text{No } H_0 \end{array} \right\} \quad (13)$$

$$\mathbf{LM}_{\text{SLM}}^* \text{SUR} = \left[\mathbf{g}(\lambda)_{|H_0} - \mathbf{I}_{\lambda, \rho, \varphi} \mathbf{I}_{\rho, \varphi}^{-1} \mathbf{g}(\rho)_{|H_0} \right]' \left[\mathbf{I}_{\lambda, \varphi} - \mathbf{I}_{\lambda, \rho, \varphi} \mathbf{I}_{\rho, \varphi}^{-1} \mathbf{I}_{\lambda, \rho, \varphi} \right]^{-1} \left[\mathbf{g}(\lambda)_{|H_0} - \mathbf{I}_{\lambda, \rho, \varphi} \mathbf{I}_{\rho, \varphi}^{-1} \mathbf{g}(\rho)_{|H_0} \right]_{\text{as}} \sim \chi^2(G)$$

$$\left. \begin{array}{l} H_0 : \rho_g = 0; \forall g \\ H_A : \text{No } H_0 \end{array} \right\} \quad (14)$$

$$\mathbf{LM}_{\text{SEM}}^* \text{SUR} = \left[\mathbf{g}(\rho)_{|H_0} - \mathbf{I}_{\lambda, \rho, \varphi} \mathbf{I}_{\lambda, \varphi}^{-1} \mathbf{g}(\lambda)_{|H_0} \right]' \left[\mathbf{I}_{\rho, \varphi} - \mathbf{I}_{\lambda, \rho, \varphi} \mathbf{I}_{\lambda, \varphi}^{-1} \mathbf{I}_{\lambda, \rho, \varphi} \right]^{-1} \left[\mathbf{g}(\rho)_{|H_0} - \mathbf{I}_{\lambda, \rho, \varphi} \mathbf{I}_{\lambda, \varphi}^{-1} \mathbf{g}(\lambda)_{|H_0} \right]_{\text{as}} \sim \chi^2(G)$$

where:

$$\begin{aligned} \mathbf{g}(\rho)_{|H_0} &= \hat{\mathbf{u}}' \left[\mathbf{I}_T \otimes \left(\boldsymbol{\Sigma}^{-1} \mathbf{E}^{\text{gg}} \right) \otimes \mathbf{W} \right] \mathbf{y} \\ \mathbf{g}(\lambda)_{|H_0} &= \hat{\mathbf{u}}' \left[\mathbf{I}_T \otimes \left(\boldsymbol{\Sigma}^{-1} \mathbf{E}^{\text{gg}} \right) \otimes \mathbf{W} \right]_{\substack{g=1, \dots, G \\ g=1, \dots, G}} \hat{\mathbf{u}} \end{aligned} \quad (15)$$

$$\mathbf{I}_{\varphi\varphi} = \begin{bmatrix} \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X} & 0 \\ 0 & \frac{\text{TR}}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1} \mathbf{E}^{\text{sr}} \right]_{i,j,s,r=1,2,\dots,G} \end{bmatrix} \quad \mathbf{I}_{\varphi\lambda} = \begin{bmatrix} \mathbf{I}_{\beta\lambda} \\ \mathbf{I}_{\sigma_{\text{gs}}\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \mathbf{E}^{\text{gg}} \otimes \mathbf{W} \right] \mathbf{X} \boldsymbol{\beta} \\ 0_{g=1,2,\dots,G} \end{bmatrix}$$

$$\mathbf{I}_{\varphi\rho} = \begin{bmatrix} \mathbf{I}_{\beta\rho} \\ \mathbf{I}_{\sigma_{\text{gs}}\rho} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \mathbf{I}_{\lambda\lambda} = \begin{bmatrix} \sigma^{\text{gs}} \left(\boldsymbol{\beta}' \mathbf{X}' \left[\mathbf{I}_T \otimes \mathbf{E}^{\text{gs}} \otimes (\mathbf{W}' \mathbf{W}) \right] \mathbf{X} \boldsymbol{\beta} + \sigma_{\text{gs}} \text{tr}(\mathbf{W}' \mathbf{W}) \right)_{g,s=1,2,\dots,G} \end{bmatrix} + (\text{Ttr} \mathbf{W}^2) \mathbf{I}_G$$

$$\mathbf{I}_{\rho\rho} = \text{Ttr}(\mathbf{W}' \mathbf{W}) \left[\left\{ \sigma^{\text{gs}} \sigma_{\text{gs}} \right\}_{g,s=1,2,\dots,G} \right] + \mathbf{I}_G \quad \mathbf{I}_{\rho\lambda} = \text{T}(\text{tr}(\mathbf{W}' \mathbf{W}) + \text{tr}(\mathbf{W} \mathbf{W})) \begin{bmatrix} \sigma^{\text{gs}} \sigma_{\text{gs}} \\ \left. \right\}_{g,s=1,2,\dots,G} \end{bmatrix}$$

$\hat{\mathbf{u}}$ is the (TGRx1) vector of residuals of the SUR model, estimated in the absence of spatial effects.

The marginal Multipliers deal with the same problem but from a different perspective given that, in the alternative hypothesis, they use the unrestricted model, $L[\varphi; \lambda; \rho]$. Specifically, for the case:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_G \quad \text{vs} \quad H_A : \text{No } H_0 \quad (16)$$

the DGP of the null hypothesis is $L_1[\varphi; \lambda] = L[\varphi; \lambda; 0]$ (the same for λ). The expression of the marginal Multiplier is:

$$\mathbf{LM}_{\text{SEM}}^{\text{SUR}}(\rho / \lambda) = \left[\mathbf{g}(\theta)_{|H_0} \right]' \left[\mathbf{I}(\theta)_{|H_0} \right]^{-1} \left[\mathbf{g}(\theta)_{|H_0} \right]_{\text{as}} \sim \chi^2(G) \quad (17)$$

where ‘ (ρ/λ) ’ means that we test a set of zero restrictions on the parameters ρ conditioned on the unrestricted ML estimation of the parameters λ and φ . The score $(g(\theta)|_{H_0})$ must be evaluated in the ML estimation of the SUR-SLM of (7):

$$g(\theta)|_{H_0} = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \lambda_g} \\ \frac{\partial l}{\partial \rho_g} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{\partial l}{\partial \rho_g} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sum_{t=1}^T \hat{u}'_{SLM,t} \left[(\Sigma^{-1} E^{gg}) \otimes \mathbf{W} \right] \hat{u}_{SLM,t} \\ 0 \end{bmatrix} \quad (18)$$

where $\hat{u}_{SLM,t}$ is the $(G \times 1)$ vector of residuals of the SLM model corresponding to the t -th cross-section. In order to obtain the value of the statistic of (17) we need the element (3,3) of the inverse of the information matrix (let us call it $I(\rho/\theta_{SLM})|_{H_0}^{-1}$, where θ_{SLM} is the vector of parameters that intervene in the SLM model: $\theta_{SLM} = [\beta, \lambda, \sigma_{ij}]'$). Using the partitioned inverse matrix we can obtain these results:

$$I(\rho/\lambda)|_{H_0}^{-1} = \left[I_{\rho,\rho} - I_{\rho,\theta_{SLM}} I_{\theta_{SLM},\theta_{SLM}}^{-1} I_{\theta_{SLM},\rho} \right]^{-1} \quad (19)$$

$$I_{\rho\rho} = T \left\{ \text{tr}(\mathbf{W}'\mathbf{W}) I_G + \begin{bmatrix} \sigma^{gs} \sigma_{gs} \\ g,s=1,2,\dots,G \end{bmatrix} \text{tr}(\mathbf{W}'\mathbf{W}) \right\}_{(G \times G)} \quad (20)$$

$$I_{\rho,\theta_{SLM}} = \begin{bmatrix} I_{\rho,\beta} & I_{\rho,\lambda} & I_{\rho,\sigma} \end{bmatrix} = \begin{bmatrix} 0 & I_{\rho,\lambda} & 0 \end{bmatrix}_{\left(G \times \left(k + G + \frac{G(G-1)}{2} \right) \right)} \quad (21)$$

$$I_{\rho,\lambda} = \left\{ \text{tr} \left[I_T \otimes (\Sigma E^{ss} \Sigma^{-1} E^{gg}) \otimes (\mathbf{W}'\mathbf{W}) + I_T \otimes (E^{gg} E^{ss}) \otimes (\mathbf{W}\mathbf{W}) \right] A^{-1} \right\}_{(G \times G)} \quad (22)$$

The discussion is similar for the parameters λ . The null hypothesis is:

$$H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_G \quad \text{vs} \quad H_A : \text{No } H_0 \quad (23)$$

in which case the DGP becomes $L_2[\varphi; \rho] = L[\varphi; 0; \rho]$. The marginal Multiplier is:

$$\mathbf{LM}_{SLM}^{SUR}(\lambda / \rho) = \left[g(\theta)|_{H_0} \right]' \left[I(\theta)|_{H_0} \right]^{-1} \left[g(\theta)|_{H_0} \right]_{as} \sim \chi^2(G) \quad (24)$$

now ‘ (λ/ρ) ’ means that we test a set of zero restrictions on the parameters λ conditioned on a unrestricted ML estimation of the parameters ρ and φ . The score $(g(\theta)|_{H_0})$ evaluated in the ML estimation of the SUR-SEM of (11) is:

$$g(\theta)|_{H_0} = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \lambda_g} \\ \frac{\partial l}{\partial \rho_g} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{\partial l}{\partial \lambda_g} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \sum_{t=1}^T \hat{\varepsilon}'_{SEM,t} \left[(\boldsymbol{\Sigma}^{-1} \mathbf{E}^{gg}) \otimes \mathbf{W} - (\boldsymbol{\Sigma}^{-1} \hat{\mathbf{Y}} \mathbf{E}^{gg}) \otimes (\mathbf{W}\mathbf{W}) \right] y_t \\ 0 \\ 0 \end{bmatrix} \quad (25)$$

where $\hat{\varepsilon}_{SEM,t}$ is the $(GR \times 1)$ vector of residuals of the SEM model corresponding to the t -th cross-section ($\hat{\varepsilon}_{SEM,t} = [\mathbf{I}_{GR} - \hat{\mathbf{Y}} \otimes \mathbf{W}] [y_t - \mathbf{X}_t \hat{\beta}]$). As before, we need the element (2,2) of the inverse of the information matrix (let us call it $I(\lambda/\theta_{SEM})|_{H_0}^{-1}$, θ_{SEM} is the vector of parameter that intervene in the SEM model: $\theta_{SEM} = [\beta, \rho, \sigma_{ij}]'$), which is equal to:

$$I(\lambda/\rho)|_{H_0}^{-1} = \left[I_{\lambda,\lambda} - I_{\lambda,\theta_{SEM}} I_{\theta_{SEM},\theta_{SEM}}^{-1} I_{\theta_{SEM},\lambda} \right]^{-1} \quad (26)$$

$$I_{\lambda\lambda} = \text{Tr}(\mathbf{W}\mathbf{W}) I_G + \left\{ \beta' \mathbf{X}' \mathbf{H}_{SEM}^{gs} \mathbf{X} \beta + \text{tr} \mathbf{H}_{SEM}^{gs} \mathbf{B}'^{-1} \boldsymbol{\Omega} \mathbf{B}^{-1} \right\}_{(G \times G)} \quad (27)$$

$$\mathbf{H}_{SEM}^{gs} = (\mathbf{I}_T \otimes \mathbf{E}^{ss} \otimes \mathbf{W})' \mathbf{B}' [\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R] \mathbf{B} (\mathbf{I} \otimes \mathbf{E}^{gg} \otimes \mathbf{W})$$

$$I_{\lambda,\theta_{SEM}} = \left[I_{\lambda,\beta} \quad I_{\lambda,\rho} \quad I_{\lambda,\sigma} \right]_{\left(G \times \left(k + G + \frac{G(G-1)}{2} \right) \right)} \quad (28)$$

$$I_{\lambda,\beta} = \left\{ \beta' \mathbf{X}' \left(\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}' \right) \mathbf{B}' \boldsymbol{\Omega}^{-1} \mathbf{B} \mathbf{X} \right\}_{(G \times k)} \quad (29)$$

$$I_{\rho,\lambda} = \left\{ \text{tr} \left\{ \boldsymbol{\Omega} \mathbf{B}'^{-1} \left(\mathbf{I}_T \otimes \mathbf{E}^{gg} \boldsymbol{\Sigma}^{-1} \otimes \mathbf{W}' \right) \mathbf{B} \left(\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W} \right) + \left(\mathbf{I}_T \otimes \mathbf{E}^{gg} \mathbf{E}^{ss} \otimes \mathbf{W}\mathbf{W} \right) \right\} \mathbf{B}^{-1} \right\}_{(G \times G)} \quad (30)$$

$$I_{\lambda,\sigma} = \left\{ \text{tr} \left[\mathbf{I}_T \otimes \left(\mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1} \right) \otimes \mathbf{I}_R \right] \mathbf{B} \left(\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W} \right) \mathbf{B}^{-1} \right\}_{\left(G \times \frac{G(G-1)}{2} \right)} \quad (31)$$

4.1- Testing for the hypothesis of common factors: SUR-SLM vs SUR-SEM models.

The hypothesis of common factors plays an important role in the specification of a cross-sectional econometric model (Burrige, 1981). In general, this test combines, in a unified framework, the information provided the marginal and the robust Multipliers⁵. This discussion is well known in the standard case and is easily transferrable to the SUR case. The ample model of the test is:

$$\left. \begin{aligned} y &= (I_T \otimes \Upsilon \otimes \mathbf{W})y + X\beta + (I_T \otimes I_G \otimes \mathbf{W})X\gamma + \varepsilon \\ \varepsilon &\sim N(0, \mathbf{\Omega}) \Rightarrow \mathbf{\Omega} = I_T \otimes \Sigma \otimes I_R \end{aligned} \right\} \quad (32)$$

As before, Υ is a $(G \times G)$ diagonal matrix of autocorrelation coefficients, β and γ are $(k \times 1)$ vectors of parameters. The hypothesis of common factors states that:

$$\left. \begin{aligned} H_0: & \left[\Upsilon \otimes I_R \right] \beta = \gamma \\ H_A: & \text{No } H_0 \end{aligned} \right\} \rightarrow \mathbf{LRCOM}_{SUR} = 2 \left[\log(H_A) - \log(H_0) \right]_{as} \sim \chi^2(Gk) \quad (33)$$

where $\log(H_A)$ is the estimated log-likelihood for the model of the alternative hypothesis (that of 32) and $\log(H_0)$ is the estimated log-likelihood corresponding to the model of the null hypothesis:

$$H_0: \left[\Upsilon \otimes I_R \right] \beta = \gamma \Rightarrow \left(\begin{bmatrix} \rho_1 & 0 & 0 & \dots & 0 \\ 0 & \rho_2 & 0 & \dots & 0 \\ 0 & 0 & \rho_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \rho_G \end{bmatrix} \otimes I_R \right) \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \dots \\ \beta_G \end{bmatrix} - \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \dots \\ \gamma_G \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} \rho_1 \beta_1 - \gamma_1 \\ \rho_2 \beta_2 - \gamma_2 \\ \rho_3 \beta_3 - \gamma_3 \\ \dots \\ \rho_G \beta_G - \gamma_G \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (34)$$

$$\begin{aligned} y &= (I_T \otimes \Upsilon \otimes \mathbf{W})y + X\beta + (I_T \otimes \Upsilon \otimes \mathbf{W})X\gamma + \varepsilon \\ &\Rightarrow \underbrace{[I_{GR} - \Upsilon \otimes \mathbf{W}]y_t - [I_{GR} - \Upsilon \otimes \mathbf{W}]X_t\beta}_{y_t - X_t\beta = u_t} = \varepsilon_t; \forall t \\ &\quad B_t u_t = \varepsilon_t \\ &\quad B_t = [I_{GR} - \Upsilon \otimes \mathbf{W}] \end{aligned}$$

Obviously, the model of the null hypothesis is the SUR-SEM of (10). If we cannot maintain this set of Gk restrictions, the model of the alternative hypothesis appears in (32). This specification will lead us to the SUR-SLM of (7) only in the case that we could not reject the additional hypothesis that vector γ is zero.

5. Some extensions of the basic SUR case.

In continuation we present some extensions of the basic spatial SUR model. In the first place, we examine the problem of the lack of constancy (among equations, between cross-sections) of the parameters of spatial dependence; then we review the question of the diagonality of the Σ matrix.

⁵ It should be remembered that what it really does is to test whether a SEM mechanism is adequate for the data (Mur and Angulo, 2006)

5.1- Testing for the constancy of the parameters of spatial dependence

The coefficients of spatial dependence of the SUR model specified in the second section are allowed to vary between equations. However, in some cases it may be important to test if these coefficients are equal for the whole set of G equations. Using another more general perspective, in some circumstances the hypothesis of interest may be that of the temporal stability of these coefficients, allowing them to they take on different values for each equation. In continuation, we discuss the first, most restrictive assumption leaving aside the case of the temporal homogeneity of the parameters⁶.

The results obtained under the assumption of constancy among equations of the coefficients of spatial dependence are equivalent to those presented in the third section, although the estimation algorithms are simpler. For example, the SUR-SARAR model of (2) now becomes:

$$\left. \begin{array}{l} \mathbf{A}\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{B}\mathbf{u} = \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim \mathbf{N}(0, \boldsymbol{\Omega}) \Rightarrow \boldsymbol{\Omega} = \mathbf{I}_T \otimes \boldsymbol{\Sigma} \otimes \mathbf{I}_R \end{array} \right\} \quad (35)$$

$$\mathbf{A} = \mathbf{I}_T \otimes [\mathbf{I}_{GR} - \boldsymbol{\Lambda} \otimes \mathbf{W}] = \mathbf{I}_T \otimes [\mathbf{I}_G \otimes (\mathbf{I}_R - \boldsymbol{\lambda}\mathbf{W})] = \mathbf{I}_T \otimes \hat{\mathbf{A}}$$

$$\mathbf{B} = \mathbf{I}_T \otimes [\mathbf{I}_{GR} - \boldsymbol{\Upsilon} \otimes \mathbf{W}] = \mathbf{I}_T \otimes [\mathbf{I}_G \otimes (\mathbf{I}_R - \boldsymbol{\rho}\mathbf{W})] = \mathbf{I}_T \otimes \hat{\mathbf{B}}$$

where $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are two $(RG \times RG)$ matrices. The number of parameters to estimate reduces to $(k+2+G(G+1)/2)$. If we use a SLM model:

$$\left. \begin{array}{l} \mathbf{A}\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim \mathbf{N}(0, \boldsymbol{\Omega}) \Rightarrow \boldsymbol{\Omega} = \mathbf{I}_T \otimes \boldsymbol{\Sigma} \otimes \mathbf{I}_R \end{array} \right\} \quad (36)$$

$$\mathbf{A} = \mathbf{I}_T \otimes [\mathbf{I}_{GR} - \boldsymbol{\Lambda} \otimes \mathbf{W}] = \mathbf{I}_T \otimes [\mathbf{I}_G \otimes (\mathbf{I}_R - \boldsymbol{\lambda}\mathbf{W})] = \mathbf{I}_T \otimes \hat{\mathbf{A}}$$

Now, the number de parameters to estimate is $(k+1+G(G+1)/2)$, the same as in the SEM case, under the assumption of constancy:

$$\left. \begin{array}{l} \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{B}\mathbf{u} = \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim \mathbf{N}(0, \boldsymbol{\Omega}) \Rightarrow \boldsymbol{\Omega} = \mathbf{I}_T \otimes \boldsymbol{\Sigma} \otimes \mathbf{I}_R \end{array} \right\} \quad (37)$$

$$\mathbf{B} = \mathbf{I}_T \otimes [\mathbf{I}_{GR} - \boldsymbol{\Upsilon} \otimes \mathbf{W}] = \mathbf{I}_T \otimes [\mathbf{I}_G \otimes (\mathbf{I}_R - \boldsymbol{\rho}\mathbf{W})] = \mathbf{I}_{TG} \otimes \hat{\mathbf{B}}$$

It is obvious that the hypothesis of homogeneity is a critical restriction that must be tested adequately. One simple solution is the likelihood ratio which compares the likelihoods of the ample and of the restricted models (that is, the model of (2) against the model of (35) in the SUR-SARAR case; the model of (7) against that of (36) in the SUR-SLM case and the model of (10) against that of (37) in the SUR-SEM case). This procedure requires the estimation of both models whereas in

⁶ The details for this case can be obtained from the authors.

the case of the Lagrange Multiplier we only need the estimation of the restricted model⁷. Other things being equal, the LM appears preferable. Below we present the details.

As said, in the SUR-SARAR case, we compare the unrestricted estimation of model (2) with the restricted version of (35). These two models are related by the following set of $2(G-1)$ restrictions:

$$H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_G \text{ and } \rho_1 = \rho_2 = \dots = \rho_G \quad \text{vs} \quad H_A : \text{No } H_0 \quad (38)$$

As it is well-known, the Lagrange Multiplier is the quadratic form of the score vector on the inverse of the information matrix, both terms evaluated under the null hypothesis. That is (Section A.IV of the Appendix):

$$\mathbf{LM}_{\text{SARAR}}^{\text{SUR}}(\lambda, \rho) = \left[\mathbf{g}(\theta)_{|H_0} \right]' \left[\mathbf{I}(\theta)_{|H_0} \right]^{-1} \left[\mathbf{g}(\theta)_{|H_0} \right]_{\text{as}} \sim \chi^2_{2(G-1)} \quad (39)$$

with:

$$\mathbf{g}(\theta)_{|H_0} = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \lambda_g} \\ \frac{\partial l}{\partial \rho_g} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix}_{|H_0} = \begin{bmatrix} \mathbf{g}(\beta)_{|H_0} \\ \mathbf{g}(\lambda)_{|H_0} \\ \mathbf{g}(\rho)_{|H_0} \\ \mathbf{g}(\sigma)_{|H_0} \end{bmatrix} = \begin{bmatrix} 0 \\ -\text{TGtr}[\hat{\mathbf{A}}^{-1}\mathbf{W}] + (\mathbf{A}y - \mathbf{X}\beta)' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}'\hat{\mathbf{B}}) \right] (\mathbf{I}_{\text{TG}} \otimes \mathbf{W}) y \\ -\text{TGtr}[\hat{\mathbf{B}}^{-1}\mathbf{W}] + (\mathbf{A}y - \mathbf{X}\beta)' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}'\mathbf{W}) \right] (\mathbf{A}y - \mathbf{X}\beta) \\ 0 \end{bmatrix} \quad (40)$$

In the case of the SUR-SLM of (7), the null hypothesis contains $(G-1)$ restrictions:

$$H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_G \quad \text{vs} \quad H_A : \text{No } H_0 \quad (41)$$

The expression of the Multiplier does not vary:

$$\mathbf{LM}_{\text{SLM}}^{\text{SUR}}(\lambda) = \left[\mathbf{g}(\theta)_{|H_0} \right]' \left[\mathbf{I}(\theta)_{|H_0} \right]^{-1} \left[\mathbf{g}(\theta)_{|H_0} \right]_{\text{as}} \sim \chi^2_{(G-1)} \quad (42)$$

with:

$$\mathbf{g}(\theta)_{|H_0} = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \lambda_g} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix}_{|H_0} = \begin{bmatrix} \mathbf{g}(\beta)_{|H_0} \\ \mathbf{g}(\lambda)_{|H_0} \\ \mathbf{g}(\sigma)_{|H_0} \end{bmatrix} = \begin{bmatrix} 0 \\ -\text{TGtr}[\hat{\mathbf{A}}^{-1}\mathbf{W}] + (\mathbf{A}y - \mathbf{X}\beta)' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R \right] (\mathbf{I}_{\text{TG}} \otimes \mathbf{W}) y \\ 0 \end{bmatrix} \quad (43)$$

Finally, in the SUR-SEM of (10) we obtain:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_G \quad \text{vs} \quad H_A : \text{No } H_0 \quad (44)$$

⁷ The Wald test is another alternative that we skip due to length restrictions.

with:

$$\mathbf{LM}_{SEM}^{SUR}(\rho) = \left[\mathbf{g}(\theta)_{|H_0} \right]' \left[\mathbf{I}(\theta)_{|H_0} \right]^{-1} \left[\mathbf{g}(\theta)_{|H_0} \right]_{as} \sim \chi^2(G-1) \quad (45)$$

where:

$$\mathbf{g}(\theta)_{|H_0} = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \rho_g} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix}_{|H_0} = \begin{bmatrix} \mathbf{g}(\beta)_{|H_0} \\ \mathbf{g}(\rho)_{|H_0} \\ \mathbf{g}(\sigma)_{|H_0} \end{bmatrix} = \begin{bmatrix} 0 \\ -\text{TGtr} \left[\hat{\mathbf{B}}^{-1} \mathbf{W} \right] + (y - \mathbf{X}\beta)' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}' \mathbf{W}) \right] (y - \mathbf{X}\beta) \\ 0 \end{bmatrix} \quad (46)$$

5.2- Testing for the diagonality of the matrix $\boldsymbol{\Sigma}$.

A characteristic of SUR models is the assumption of linear dependence between the random terms of the G equations, which explains the importance of the hypothesis of non-diagonality of the $\boldsymbol{\Sigma}$ matrix. This problem is well known in the literature where we can find different proposals. Among them, the Likelihood Ratio and the Lagrange Multiplier tests (Breusch and Pagan, 1980) fit properly in our framework. The only aspect that we should consider is that these diagonality tests will be used in models where there exists a certain spatial structure. For example, in the case of the SUR-SARAR of (2):

$$\left. \begin{array}{l} \mathbf{A}\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{B}\mathbf{u} = \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim \mathbf{N}(0, \boldsymbol{\Omega}) \Rightarrow \boldsymbol{\Omega} = \mathbf{I}_T \otimes \boldsymbol{\Sigma} \otimes \mathbf{I}_R \end{array} \right\} \quad (47)$$

matrix $\boldsymbol{\Sigma}$ refers to the correlation structure that exists between the errors of the G equations, once the spatial dependence that intervenes in the specification has been filtered. Then, we obtain the test in the usual way:

$$\left. \begin{array}{l} H_0 : \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_1^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_G^2 \end{bmatrix} \\ H_A : \quad \quad \quad \text{No } H_0 \end{array} \right\} \rightarrow \lambda_{LM} = \text{TR} \boldsymbol{\Sigma}_{g=1}^G \sum_{s=1}^{g-1} r_{gs}^2 \sim \chi^2_{as}(G(G-1)/2) \quad (48)$$

with:

$$r_{gs} = \frac{\sum_{t=1}^T \sum_{r=1}^R (\hat{\varepsilon}_{trg} - \hat{\varepsilon}_g) (\hat{\varepsilon}_{trs} - \hat{\varepsilon}_s)}{\sqrt{\sum_{t=1}^T \sum_{r=1}^R (\hat{\varepsilon}_{trg} - \hat{\varepsilon}_g)^2} \sqrt{\sum_{t=1}^T \sum_{r=1}^R (\hat{\varepsilon}_{trs} - \hat{\varepsilon}_s)^2}} \quad (49)$$

r_{gs} is the correlation coefficient obtained between the residuals of equations g and s. Breusch and Pagan (1980) use the LS residuals because they discuss the assumption of diagonality in a static

model, but this is not our case. In the framework of (47), and under the null hypothesis of diagonality, we have a series of ML residuals obtained from the estimation of a spatial model (SARAR, SLM or SEM) for each of the G equations.

6. A Monte Carlo study on the problem of specifying spatial SUR models.

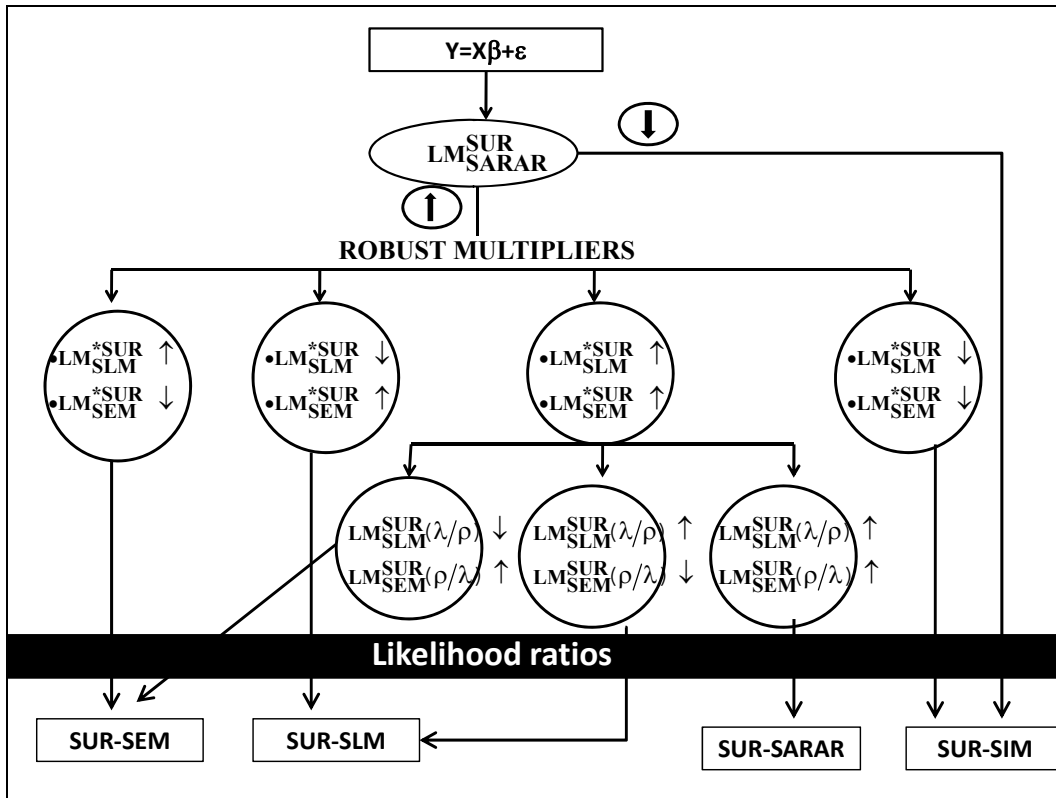
The results of the previous section should help us to improve the specification of a given spatial SUR model. We are confident in these techniques because, in a previous study (Mur and López, 2010), simpler versions of them worked reasonably well. We are going to focus the discussion in another aspect of the problem; specifically, in the question of selecting the most adequate spatial structure for the data at hand. In the remaining of this section, we are going to compare the performance of two well-known selection strategies: a General-to-Specific, *Gets* in short, or a Specific-to-General, *Stge*, approach. This problem has received a considerable attention in mainstream econometrics where there exists a huge amount of published work (see, for example, Charemza and Deadman, 1997, Campos et al, 2005, or Lütkepohl, 2007). Recently, in a context of spatial modeling, this debate has been raised by Florax et al (2003, see also Florax et al, 2006, Hendry, 2006, and Mur and Angulo, 2009).

The evidence is not totally conclusive: whereas the *Gets* approach seems to be more robust to severe misspecifications of the model, the *Stge* approach is simpler and more efficient when the model is reasonably well-specified. Our purpose is to extend this discussion to a spatial SUR model using the tests developed in the previous sections. First we briefly present the two strategies; section 6.2 focuses on the design of the Monte Carlo and in section 6.3 the main results of this Monte Carlo are discussed.

6.1- Gets vs Stge: Main characteristics.

In Figure 1 we present a sketch of a *Stge* strategy to solve the problem of selecting the most appropriate specification for a SUR model (other combinations of the statistics, in a *Stge* framework, are possible; we present our favorite alternative).

FIGURE 1: A Stge algorithm to spatial SUR model selection* .



* \uparrow Reject Null Hypothesis ; \downarrow Non Reject Null Hypothesis

The starting point is the simplest admissible model which, in our case, will be the SUR model with no spatial effects (Spatially Independent Model, SIM from now on). LM_{SARAR}^{SUR} is the adequate statistic at this stage. The SIM model will be chosen if the null cannot be rejected and the process finishes at this point. On the contrary, we will use the robust Multipliers to select the SUR model with spatial effect that corresponds to the alternative hypothesis. Four different situations can be considered:

- (a)- LM_{SEM}^{SUR} is significant but LM_{SLM}^{SUR} is not statistically significant. The model appears to be a SEM. We confirm this identification by means of the corresponding likelihood ratio.
- (b)- LM_{SLM}^{SUR} is significant but LM_{SEM}^{SUR} is not statistically significant. The model appears to be a SLM. As before, we confirm the selection by means of the corresponding likelihood ratio.
- (c)- Both robust Multipliers reject their respective null hypotheses. We use the marginal Multipliers:
 - (c.1)- $LM_{SLM}^{SUR}(\lambda/\rho)$ and $LM_{SEM}^{SUR}(\rho/\lambda)$ are statistically significant; we select a SARAR model.

(c.2)- Only $\mathbf{LM}_{\text{SEM}}^{\text{SUR}}(\rho / \lambda)$ is statistically significant; we select a SEM model.

(c.3)- Only $\mathbf{LM}_{\text{SLM}}^{\text{SUR}}(\lambda / \rho)$ is statistically significant; we select a SLM model.

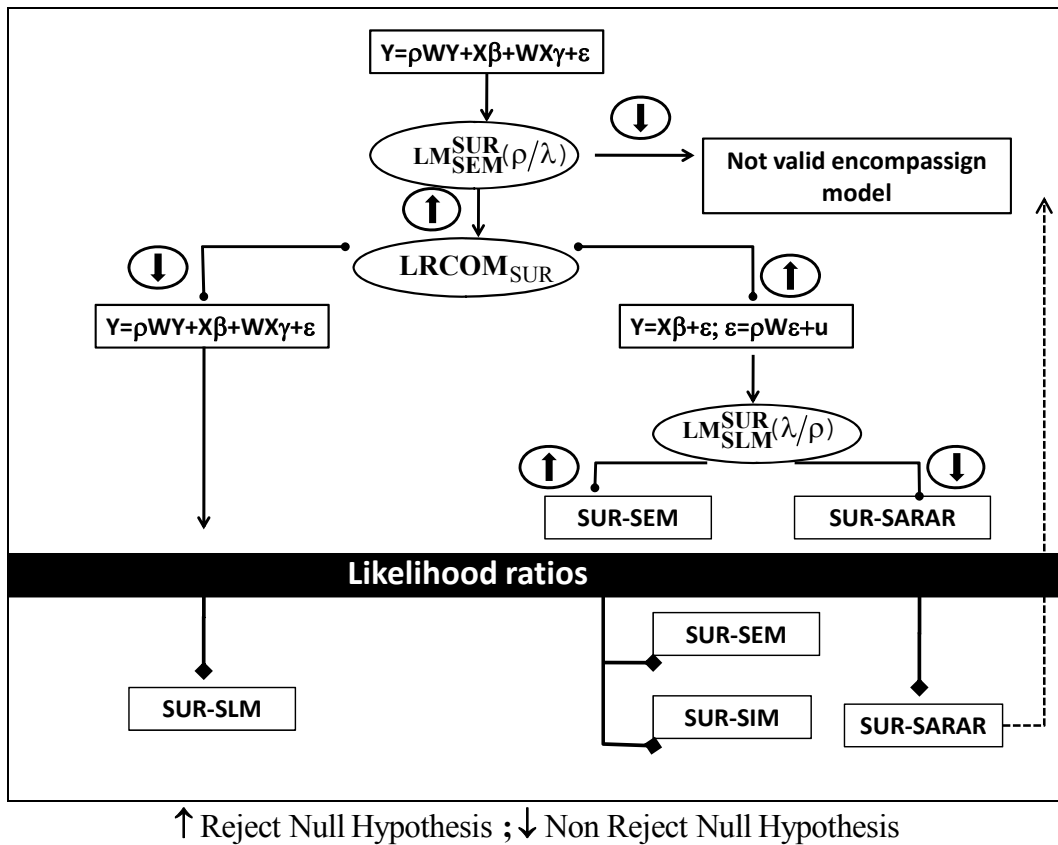
(d)- None of the two null hypotheses can be rejected with the robust Multipliers. We conclude that it is a SIM model.

Figure 2 describes the functioning of a Gets algorithm with the statistics that we have developed (other combinations are possible). According to Hendry (1980) the idea is to start with a very general model in order to avoid unnecessary restrictions. This model must be consistent with the data (i.e., the random term must behave as expected, and all other hypotheses should also be fulfilled). In our case, let us start with the ‘*the autoregressive distributed lag model of the first order*’ ADL introduced by Bivand (1984, p27):

$$\left. \begin{aligned} y &= (\mathbf{I}_T \otimes \Upsilon \otimes \mathbf{W})y + X\beta + (\mathbf{I}_T \otimes \mathbf{I}_G \otimes \mathbf{W})X\gamma + \varepsilon \\ \varepsilon &\sim \mathbf{N}(0, \mathbf{\Omega}) \Rightarrow \mathbf{\Omega} = \mathbf{I}_T \otimes \Sigma \otimes \mathbf{I}_R \end{aligned} \right\} \quad (50)$$

As indicated, this model must be consistent with the data which means that the error term of (50) should be spatially independent. That is the point of the $\mathbf{LM}_{\text{SEM}}^{\text{SUR}}(\rho/\lambda)$ test. If we cannot maintain this assumption, the model of (50) is not valid as ‘*encompassing model*’. If, on the contrary, we cannot reject the null hypothesis, the Gets algorithm continues with the simplification process depicted in Figure 2.

FIGURE 2 A Gets algorithm to spatial SUR model selection



The Common Factor test plays a very important role in the Gets algorithm in order to discriminate between SEM and SLM models. We have two options:

- (i)- If we reject the Common Factor hypothesis, the evidence is in favor of a SLM model. The likelihood ratio should confirm this identification.
- (ii)- If we cannot reject the Common Factor hypothesis, there is a SEM structure in the data, pure or combined with a SLM component. We will solve this point using the marginal Multiplier, $LM_{SLM}^{SUR}(\lambda / \rho)$, in the following sense:

- (i)- A pure SEM model emerges when we cannot reject the null hypothesis.
- (ii)- A SARAR model is advisable when the previous null hypothesis is rejected.

The last result is an indication against the encompassing model of (50):

6.2- Design of the Monte Carlo.

The purpose of this exercise is to compare the performance of the two strategies of model selection in a SUR context. We are going to simulate in the Data Generating Process up to four processes: a SIM, a SEM, a SLM and a SARAR. Success of a strategy, Stge or Gets, means highest probability of selecting the right model.

We have included only one regressor, plus a constant term equal to 2, in each equation of the DGP. The regressor has been obtained from a $N(0,1)$ distribution and each regressor enters into the

equation with a coefficient of 3. This assures that, in the absence of spatial effects, the R^2 statistic of the Least Square (LS) regression of each equation will take, on average, a value of 0.9. The SUR system is made by 3 equations ($G=3$), 3 cross-sections ($T=3$) and 49 spatial units ($R=49$). Overall, the results are quite robust to these values. Moreover, we have used a (7x7) regular lattice assuming rook contiguities in order to obtain the weighting W matrix, which has been row-standardized.

The SUR structure corresponds to the $\Sigma_\varepsilon = \{\sigma_{ij} = 1, \text{ if } i=j; \sigma_{ij} = c, \text{ if } i \neq j\}$ matrix in (2) for which we have used two specifications according to the values of c : a medium level of cross-error dependence, $c=0.5$, and a high level, $c=0.9$. The values used in the spatial correlation coefficients, ρ and λ , are 0.1, 0.3, 0.5 and 0.9.

We study the behavior of the two specification strategies under the following situations:

- (i)- Ideal conditions (normality, homoskedasticity, ...), though we do not know which DGP (SIM, SEM, SLM or SARAR) has generated the data.
- (ii)- Non-normality in the error terms. We consider two possibilities: a Wishart distribution with covariance matrix $\Sigma_{0,9}$ and 1 degree of freedom and a Multinomial distribution $MN(10; 0.5, 0.3, 0.2)$. The tails of the Wishart distribution are more dense than in the normal case whereas the Multinomial is an extreme case of non normality and negative correlation.
- (iii)- Heteroskedasticity. We allow the variance of random term of each equation to vary across time, maintaining constant the correlations of the SUR structure. That is, the specification of the matrix Ω now becomes: $\Omega = \Delta_T \otimes \Sigma_\varepsilon \otimes I_R$ being Δ_T a diagonal matrix. We have tried two different mechanisms of heteroskedasticity: (i)- A random structure which means that the elements of matrix Δ_T are obtained, in each trial, from a $U(0,1)$ distribution; (ii)- An increasing pattern of heteroskedasticity reflected in the values the diagonal matrix Δ_T as: $\Delta_T = [\delta_{11} = 1; \delta_{jj} = \delta_{j-1j-1}(1 + \pi); j = 2, \dots, T; \pi=0.1]$; this means that the variance increases by a 10% in time period.
- (iv)- Misspecification of the weighting matrix. In this case, the weighting matrix of the DGP is of the queen type whereas the tests are solved using a rook type matrix.
- (v)- Endogeneity. We introduce some dependence between the regressor and the error term of each equation using a linear or a non-linear relationship:

(v.a) $X = \alpha_1 \varepsilon + \alpha_2 \eta$; where $\alpha_1 + \alpha_2 = 1$, $\alpha_1; \alpha_2 \geq 0$ and $\eta \sim N(0,1)$.

(v.b) $X = \exp(\alpha_1 \varepsilon) + \exp(\alpha_2 \eta)$; where $\alpha_1 + \alpha_2 = 1$, $\alpha_1; \alpha_2 \geq 0$ and $\eta \sim N(0,1)$.

Finally, each experiment has been repeated 1000 times.

6.3- Main results

The data of Table 1 are the percentages that the two strategies correctly select the process of the DGP, simulated under ideal conditions. The first column identifies the DGP used in the simulation, the second column the values of the coefficients of spatial dependence and the degree of correlation between the equation; the following two blocks contain the percentages corresponding to a Stge strategy and to the Gets approach.

Table 1: Stge vs Gets under ideal conditions

				Percentages of correct identification of the DGP							
				Stge				Gets			
	c	λ	ρ	SIM	SLM	SEM	SARAR	SIM	SLM	SEM	SARAR
SIM	0.0	--	--	0.965	0.010	0.022	0.003	0.923	0.018	0.051	0.008
SLM	0.3	0.1	--	0.188	0.762	0.003	0.047	0.130	0.750	0.021	0.099
	0.3	0.3	--	0.000	0.949	0.000	0.051	0.000	0.913	0.000	0.087
	0.3	0.5	--	0.000	0.955	0.000	0.045	0.000	0.914	0.000	0.086
	0.3	0.9	--	0.000	0.949	0.000	0.051	0.000	0.933	0.000	0.067
SEM	0.3	--	0.1	0.884	0.031	0.079	0.004	0.792	0.030	0.155	0.023
	0.3	--	0.3	0.073	0.005	0.894	0.024	0.020	0.004	0.888	0.088
	0.3	--	0.5	0.000	0.000	0.974	0.026	0.000	0.000	0.892	0.108
	0.3	--	0.9	0.007	0.000	0.956	0.037	0.000	0.000	0.841	0.159
SARAR	0.3	0.1	0.1	0.118	0.732	0.064	0.083	0.044	0.692	0.110	0.154
	0.3	0.3	0.3	0.000	0.064	0.000	0.936	0.000	0.059	0.000	0.941
	0.3	0.5	0.5	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
	0.3	0.9	0.9	0.005	0.000	0.000	0.995	0.000	0.000	0.000	1.000
SLM	0.9	0.1	--	0.000	0.951	0.000	0.049	0.006	0.893	0.000	0.101
	0.9	0.3	--	0.000	0.951	0.000	0.049	0.000	0.910	0.000	0.090
	0.9	0.5	--	0.000	0.956	0.000	0.044	0.000	0.911	0.000	0.089
	0.9	0.9	--	0.000	0.952	0.000	0.048	0.000	0.933	0.000	0.067
SEM	0.9	--	0.1	0.887	0.019	0.075	0.004	0.973	0.003	0.021	0.003
	0.9	--	0.3	0.065	0.000	0.887	0.045	0.441	0.000	0.500	0.059
	0.9	--	0.5	0.000	0.000	0.954	0.046	0.010	0.000	0.864	0.126
	0.9	--	0.9	0.008	0.000	0.963	0.029	0.000	0.000	0.874	0.126
SARAR	0.9	0.1	0.1	0.000	0.837	0.000	0.163	0.002	0.774	0.000	0.224
	0.9	0.3	0.3	0.000	0.023	0.000	0.977	0.000	0.019	0.000	0.981
	0.9	0.5	0.5	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000

In general terms, we can say that both strategies perform reasonably. The overall percentage of correct identifications 81.3% for the Stge approach and 77.7% for the Gets approach. Both algorithms perform better with SLM models (with percentages of 92.8% and 89.5) than with SEM (72.3% and 62.9% respectively). The Gets only appears slightly superior if the DGP is SARAR. Moreover, the correlation between the equations appears to have a minimal impact on the performance of both algorithms, which are more sensible to the intensity of the spatial dependence, especially for SEM processes.

Tables 2a, 2b show the results corresponding to the case of non-normality. In the first case, we used a Wishart distribution and a Multinomial in the second table.

Table 2a: Stge vs Gets with Non-normal errors. Wishart distribution

		Percentages of correct identification of the DGP										
		Stge						Gets				
		c	λ	ρ	SIM	SLM	SEM	SARAR	SIM	SLM	SEM	SARAR
SIM		0.9	0.0	0.0	0.968	0.010	0.011	0.004	0.983	0.009	0.004	0.004
SLM	0.9	0.1	--	0.133	0.816	0.002	0.038	0.183	0.722	0.004	0.091	
	0.9	0.3	--	0.005	0.952	0.000	0.043	0.000	0.910	0.000	0.090	
	0.9	0.5	--	0.002	0.961	0.000	0.037	0.000	0.929	0.000	0.071	
	0.9	0.9	--	0.000	0.958	0.000	0.042	0.000	0.942	0.000	0.058	
SEM	0.9	--	0.1	0.885	0.016	0.077	0.010	0.939	0.008	0.048	0.005	
	0.9	--	0.3	0.157	0.003	0.803	0.032	0.190	0.001	0.726	0.083	
	0.9	--	0.5	0.061	0.000	0.907	0.032	0.001	0.000	0.879	0.120	
	0.9	--	0.9	0.004	0.000	0.960	0.036	0.000	0.000	0.874	0.126	
SARAR	0.9	0.1	0.1	0.105	0.720	0.019	0.139	0.093	0.694	0.012	0.201	
	0.9	0.3	0.3	0.010	0.041	0.000	0.949	0.000	0.034	0.000	0.966	
	0.9	0.5	0.5	0.007	0.000	0.000	0.993	0.000	0.000	0.000	1.000	
	0.9	0.9	0.9	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	

The consequences of using such non-normal distributions are dramatic. Overall, the percentage of correct selections decreases to 63.5% for the Stge strategy and 60.5% in the case of Gets. The SEM processes are very bad identified in both cases, with a ridiculous percentage of right decisions that almost zero in the case of the Multinomial. On the contrary, SARAR processes are reasonably well treated with percentages above 80% for the two algorithms and with both distributions (Wishart and Multinomial).

Table 2b: Stge vs Gets with Non-normal errors. Multinomial distribution.**Percentages of correct identification of the DGP**

				Stge				Gets			
	c	λ	ρ	SIM	SLM	SEM	SARAR	SIM	SLM	SEM	SARAR
SIM	0.9	0.0	0.0	0.984	0.008	0.000	0.008	0.949	0.003	0.010	0.038
SLM	0.9	0.1	--	0.004	0.699	0.000	0.297	0.001	0.633	0.000	0.366
	0.9	0.3	--	0.014	0.610	0.000	0.376	0.000	0.456	0.000	0.544
	0.9	0.5	--	0.000	0.498	0.000	0.502	0.000	0.235	0.000	0.765
	0.9	0.9	--	0.000	0.185	0.000	0.815	0.000	0.089	0.000	0.911
SEM	0.9	--	0.1	0.945	0.018	0.000	0.033	0.944	0.007	0.007	0.042
	0.9	--	0.3	0.489	0.069	0.010	0.430	0.870	0.024	0.004	0.102
	0.9	--	0.5	0.151	0.007	0.005	0.837	0.714	0.002	0.000	0.284
	0.9	--	0.9	0.000	0.000	0.046	0.954	0.295	0.000	0.000	0.705
SARAR	0.9	0.1	0.1	0.011	0.545	0.000	0.444	0.000	0.476	0.000	0.524
	0.9	0.3	0.3	0.002	0.101	0.000	0.897	0.000	0.004	0.000	0.996
	0.9	0.5	0.5	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.000	0.008	0.000	0.992	0.000	0.001	0.000	0.999

The results corresponding to heteroskedasticity appear in Tables 3a-3b. The performance of both strategies worsens only slightly in relation to the results observed under ideal conditions. Now the percentage of correct identifications is 81.8% for the Stge algorithm and 75.6% for the Gets. Once more, the worse results correspond to SEM processes where these percentages are 71.9% and 53.2% respectively. The pattern of heteroskedasticity, random or increasing along time, has a minimal impact on the aggregated scores.

Table 3a: Stge vs Gets with heteroskedasticity. $\Omega = \Delta_T \otimes \Sigma_\varepsilon \otimes I_R$. Random pattern

				Stge				Gets			
	c	λ	ρ	SIM	SLM	SEM	SARAR	SIM	SLM	SEM	SARAR
SIM	0.9	0.0	0.0	0.969	0.008	0.018	0.005	0.912	0.026	0.053	0.009
SLM	0.9	0.1	--	0.000	0.960	0.000	0.040	0.000	0.905	0.000	0.095
	0.9	0.3	--	0.000	0.958	0.000	0.042	0.000	0.924	0.000	0.076
	0.9	0.5	--	0.000	0.953	0.000	0.047	0.000	0.923	0.000	0.077
	0.9	0.9	--	0.000	0.949	0.000	0.051	0.000	0.939	0.000	0.061
SEM	0.9	--	0.1	0.874	0.030	0.083	0.008	0.979	0.006	0.015	0.000
	0.9	--	0.3	0.060	0.001	0.890	0.046	0.432	0.000	0.512	0.056
	0.9	--	0.5	0.000	0.000	0.958	0.042	0.012	0.000	0.871	0.117
	0.9	--	0.9	0.013	0.000	0.939	0.048	0.000	0.000	0.847	0.153
SARAR	0.9	0.1	0.1	0.000	0.832	0.000	0.168	0.000	0.779	0.000	0.221
	0.9	0.3	0.3	0.000	0.032	0.000	0.968	0.000	0.023	0.000	0.977
	0.9	0.5	0.5	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.001	0.000	0.000	1.000	0.000	0.000	0.000	1.000

Table 3b: Stge vs Gets with heteroskedasticity. $\Omega = \Delta_T \otimes \Sigma_\varepsilon \otimes I_R$. Increasing pattern

				Stge				Gets			
	c	λ	ρ	SIM	SLM	SEM	SARAR	SIM	SLM	SEM	SARAR
SIM	0.9	0.0	0.0	0.946	0.010	0.039	0.004	0.881	0.024	0.083	0.012
SLM	0.9	0.1	--	0.000	0.907	0.000	0.093	0.000	0.846	0.000	0.154
	0.9	0.3	--	0.000	0.908	0.000	0.092	0.000	0.868	0.000	0.132
	0.9	0.5	--	0.000	0.900	0.000	0.100	0.000	0.866	0.000	0.134
	0.9	0.9	--	0.000	0.897	0.000	0.103	0.000	0.884	0.000	0.116
SEM	0.9	--	0.1	0.845	0.030	0.111	0.012	0.960	0.006	0.026	0.008
	0.9	--	0.3	0.082	0.001	0.865	0.050	0.406	0.001	0.510	0.083
	0.9	--	0.5	0.000	0.000	0.965	0.035	0.021	0.000	0.825	0.154
	0.9	--	0.9	0.010	0.000	0.944	0.046	0.000	0.000	0.810	0.190
SARAR	0.9	0.1	0.1	0.000	0.799	0.000	0.201	0.000	0.743	0.000	0.257
	0.9	0.3	0.3	0.000	0.041	0.000	0.959	0.000	0.038	0.000	0.962
	0.9	0.5	0.5	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.001	0.000	0.000	0.999	0.000	0.000	0.000	1.000

Table 4 focuses on the impact of a misspecified weighting matrix (we have defined less connections than the real ones). The consequences of this systematic under-specification are very severe on both approaches. When the DGP is of the SLM type, the two strategies obtain very poor scores, well below the 50% of correct identifications with a minimum 5.5% in the case of the Gets algorithm. The situation improves slightly if the process is of the SEM type (57.5% and 10.5% respectively), and changes drastically for SARAR processes where the Gets algorithm attains an 80.0% of correct identification whereas the Stge percentage is only 15.7%.

Table 4: Stge versus Gets under misspecification of W.

		Percentages of correct identification of the DGP									
		Stge				Gets					
	c	λ	ρ	SIM	SLM	SEM	SARAR	SIM	SLM	SEM	SARAR
SIM	0.9	0.0	0.0	0.965	0.010	0.022	0.003	0.923	0.018	0.051	0.008
SLM	0.9	0.1	--	0.001	0.945	0.000	0.053	0.122	0.219	0.000	0.659
	0.9	0.3	--	0.000	0.768	0.000	0.232	0.000	0.000	0.000	1.000
	0.9	0.5	--	0.000	0.179	0.016	0.802	0.000	0.000	0.000	1.000
	0.9	0.9	--	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
SEM	0.9	--	0.1	0.916	0.026	0.044	0.006	0.982	0.007	0.008	0.003
	0.9	--	0.3	0.401	0.020	0.517	0.047	0.800	0.002	0.144	0.054
	0.9	--	0.5	0.004	0.000	0.919	0.076	0.200	0.000	0.268	0.532
	0.9	--	0.9	0.014	0.000	0.819	0.167	0.000	0.000	0.000	1.000
SARAR	0.9	0.1	0.1	0.002	0.919	0.003	0.070	0.101	0.215	0.000	0.684
	0.9	0.3	0.3	0.000	0.860	0.000	0.138	0.000	0.000	0.000	1.000
	0.9	0.5	0.5	0.001	0.837	0.004	0.144	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.000	0.726	0.000	0.274	0.000	0.483	0.000	0.517

Finally, Tables 5a and 5b contain the details for the case of endogeneity, of a linear type in Tables 5a and of a nonlinear type in Tables 5b. The behavior of both strategies sharply deteriorates as the endogeneity becomes stronger. In this case, there is a strong tendency for the two algorithms to identify mixed SARAR processes. This is particularly evident in the case of SEM processes with a percentage of correct identifications almost zero in the linear and in the nonlinear case. The Stge algorithm appears to be more robust when the process of the SLM type whereas the Gets approach attains better results in the case of SARAR processes.

Table 5a: Stge vs Gets under endogeneity. Linear case: $X = \alpha_1 \varepsilon + \alpha_2 \eta$

				Percentages of correct identification of the DGP							
				Stge				Gets			
	c	λ	ρ	SIM	SLM	SEM	SARAR	SIM	SLM	SEM	SARAR
$\alpha_1 = 0.2 \quad \alpha_2 = 0.8$											
SIM	0.9	0.0	0.0	0.962	0.019	0.018	0.001	0.936	0.000	0.049	0.015
SLM	0.9	0.1	--	0.005	0.970	0.000	0.025	0.833	0.000	0.000	0.167
	0.9	0.3	--	0.000	0.970	0.000	0.030	0.000	0.000	0.000	1.000
	0.9	0.5	0.0	0.000	0.962	0.000	0.038	0.000	0.000	0.000	1.000
	0.9	0.9	0.0	0.002	0.956	0.000	0.042	0.000	0.000	0.000	1.000
SEM	0.9	--	0.1	0.879	0.024	0.052	0.008	1.000	0.000	0.000	0.000
	0.9	--	0.3	0.211	0.003	0.724	0.058	0.668	0.000	0.331	0.001
	0.9	0.0	0.5	0.011	0.000	0.931	0.058	0.307	0.000	0.660	0.033
	0.9	0.0	0.9	0.012	0.000	0.912	0.076	0.000	0.000	0.320	0.680
SARAR	0.9	0.1	0.1	0.006	0.868	0.000	0.126	0.725	0.000	0.000	0.275
	0.9	0.3	0.3	0.001	0.044	0.000	0.955	0.000	0.000	0.000	1.000
	0.9	0.5	0.5	0.003	0.000	0.000	0.997	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.962	0.019	0.018	0.001	0.936	0.000	0.049	0.015
$\alpha_1 = 0.5 \quad \alpha_2 = 0.5$											
SIM	0.9	0.0	0.0	0.965	0.018	0.010	0.006	0.932	0.000	0.042	0.026
SLM	0.9	0.1	--	0.173	0.808	0.000	0.019	0.712	0.000	0.003	0.285
	0.9	0.3	--	0.000	0.986	0.000	0.014	0.000	0.000	0.000	1.000
	0.9	0.5	--	0.000	0.985	0.000	0.015	0.000	0.000	0.000	1.000
	0.9	0.9	--	0.001	0.963	0.000	0.036	0.000	0.000	0.000	1.000
SEM	0.9	--	0.1	0.898	0.043	0.034	0.006	0.549	0.000	0.428	0.023
	0.9	--	0.3	0.333	0.036	0.304	0.327	0.381	0.000	0.535	0.084
	0.9	--	0.5	0.002	0.000	0.547	0.451	0.083	0.000	0.564	0.353
	0.9	--	0.9	0.040	0.000	0.321	0.639	0.000	0.000	0.092	0.908
SARAR	0.9	0.1	0.1	0.071	0.836	0.000	0.093	0.583	0.000	0.002	0.415
	0.9	0.3	0.3	0.000	0.092	0.000	0.908	0.000	0.000	0.000	1.000
	0.9	0.5	0.5	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
$\alpha_1 = 0.8 \quad \alpha_2 = 0.2$											
SIM	0.9	0.0	0.0	0.962	0.018	0.018	0.002	0.932	0.000	0.045	0.023
SLM	0.9	0.1	--	0.000	0.952	0.000	0.048	0.000	0.000	0.000	1.000
	0.9	0.3	--	0.000	0.941	0.000	0.059	0.000	0.000	0.000	1.000
	0.9	0.5	--	0.000	0.962	0.000	0.038	0.000	0.000	0.000	1.000
	0.9	0.9	--	0.000	0.954	0.000	0.046	0.000	0.000	0.000	1.000
SEM	0.9	--	0.1	0.102	0.802	0.008	0.085	0.166	0.000	0.007	0.827
	0.9	--	0.3	0.000	0.264	0.000	0.736	0.000	0.000	0.000	1.000
	0.9	--	0.5	0.000	0.002	0.000	0.998	0.000	0.000	0.000	1.000
	0.9	--	0.9	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
SARAR	0.9	0.1	0.1	0.000	0.922	0.000	0.078	0.000	0.000	0.000	1.000
	0.9	0.3	0.3	0.000	0.259	0.000	0.741	0.000	0.000	0.000	1.000
	0.9	0.5	0.5	0.000	0.002	0.000	0.998	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000

Table 5b: Stge vs Gets under endogeneity. Nonlinear case: $X=\exp(\alpha_1\varepsilon)+\exp(\alpha_2\eta)$

				Percentages of correct identification of the DGP							
				Stge				Gets			
c	λ	ρ		SIM	SLM	SEM	SARAR	SIM	SLM	SEM	SARAR
$\alpha_1=0.2$		$\alpha_2=0.8$									
SIM	0.9	0.0	0.0	0.967	0.011	0.019	0.003	0.928	0.000	0.057	0.015
SLM	0.9	0.1	--	0.000	0.969	0.000	0.031	0.023	0.000	0.000	0.977
	0.9	0.3	--	0.000	0.967	0.000	0.033	0.000	0.000	0.000	1.000
	0.9	0.5	0.0	0.000	0.959	0.000	0.041	0.000	0.000	0.000	1.000
	0.9	0.9	0.0	0.000	0.956	0.000	0.044	0.000	0.000	0.000	1.000
SEM	0.9	--	0.1	0.887	0.027	0.063	0.006	1.000	0.000	0.000	0.000
	0.9	--	0.3	0.133	0.002	0.820	0.044	0.918	0.000	0.075	0.007
	0.9	0.0	0.5	0.000	0.000	0.941	0.059	0.241	0.000	0.694	0.065
	0.9	0.0	0.9	0.016	0.000	0.927	0.057	0.000	0.000	0.522	0.478
SARAR	0.9	0.1	0.1	0.000	0.850	0.000	0.150	0.019	0.000	0.000	0.981
	0.9	0.3	0.3	0.000	0.044	0.000	0.956	0.000	0.000	0.000	1.000
	0.9	0.5	0.5	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
$\alpha_1=0.5$		$\alpha_2=0.5$									
SIM	0.9	0.0	0.0	0.964	0.021	0.009	0.006	0.929	0.000	0.039	0.032
SLM	0.9	0.1	--	0.067	0.908	0.000	0.025	0.864	0.000	0.000	0.136
	0.9	0.3	--	0.000	0.987	0.000	0.013	0.000	0.000	0.000	1.000
	0.9	0.5	--	0.000	0.981	0.000	0.019	0.000	0.000	0.000	1.000
	0.9	0.9	--	0.002	0.964	0.000	0.034	0.000	0.000	0.000	1.000
SEM	0.9	--	0.1	0.916	0.031	0.027	0.012	0.944	0.000	0.055	0.001
	0.9	--	0.3	0.371	0.021	0.335	0.273	0.839	0.000	0.142	0.019
	0.9	--	0.5	0.006	0.000	0.629	0.365	0.364	0.000	0.435	0.201
	0.9	--	0.9	0.030	0.000	0.403	0.567	0.000	0.000	0.158	0.842
SARAR	0.9	0.1	0.1	0.042	0.858	0.000	0.100	0.755	0.000	0.000	0.245
	0.9	0.3	0.3	0.000	0.076	0.000	0.924	0.000	0.000	0.000	1.000
	0.9	0.5	0.5	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
$\alpha_1=0.8$		$\alpha_2=0.2$									
SIM	0.9	0.0	0.0	0.969	0.021	0.009	0.001	0.927	0.000	0.046	0.027
SLM	0.9	0.1	--	0.000	0.957	0.000	0.043	0.029	0.000	0.000	0.971
	0.9	0.3	--	0.000	0.971	0.000	0.029	0.000	0.000	0.000	1.000
	0.9	0.5	--	0.000	0.968	0.000	0.032	0.000	0.000	0.000	1.000
	0.9	0.9	--	0.000	0.962	0.000	0.038	0.000	0.000	0.000	1.000
SEM	0.9	--	0.1	0.856	0.081	0.042	0.020	0.994	0.000	0.002	0.004
	0.9	--	0.3	0.017	0.104	0.092	0.787	0.327	0.000	0.044	0.629
	0.9	--	0.5	0.000	0.000	0.001	0.999	0.000	0.000	0.001	0.999
	0.9	--	0.9	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
SARAR	0.9	0.1	0.1	0.000	0.905	0.000	0.095	0.002	0.000	0.000	0.998
	0.9	0.3	0.3	0.000	0.124	0.000	0.876	0.000	0.000	0.000	1.000
	0.9	0.5	0.5	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
	0.9	0.9	0.9	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000

6. Final Conclusions

SUR models approach is a very popular technique to deal simultaneously with multidimensional data, which requires few assumptions and computation efforts. However, as far as we know, since the seminal work of Anselin (1988a and b), very little has been said about combining SUR models with spatial processes. We have tried to fill this gap. Specifically, we have developed a collection of tests, based on the principle of the Lagrange Multiplier, that are efficient in detecting the presence of spatial elements in SUR models.

According to the results of the simulation, a traditional Specific-to-General approach is slightly preferable to the inverse General-to-Specific procedure. Both algorithms of model selection seem to work reasonable well under ideal conditions. The Stge procedure tends to select more parsimonious models whereas the Gets algorithm shows a high propensity towards SARAR models. However, the two approaches are very sensitive to anomalies in the DGP. Misspecification of the weighting matrix and (severe) departures from the assumption of normality affect to a great extent the behavior of both algorithms. Endogeneity and heteroskedasticity have more impact on the Gets procedure.

References

- Anselin, L. (1988a): *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.
- Anselin L. (1988b): A test for Spatial Autocorrelation in Seemingly Unrelated Regressions, *Economics Letters*, **28**, 335-341.
- Anselin L. (1990): Spatial dependence and spatial structural instability in applied regression analysis. *Journal of Regional Science* **30**, 185–207.
- Anselin, L., A. Bera, R. Florax and M. Yoon, (1996): Simple Diagnostic Tests for Spatial Dependence. *Regional Science and Urban Economics*, **26**, 77-104.
- Anselin L, LeGallo, J. and H. Jayet (2007): Spatial panel econometrics. In Matyas L. and Sevestre P. (Eds.): *The Econometrics of Panel Data*. Dordrecht: Kluwer Academic Publishers.
- Arora, S. and M. Brown (1977): Alternative Approaches to Spatial Autocorrelation: An Improvement over Current Practice. *International Regional Science Review*, **2**, 67-78.
- Badinger H., W. Müller and G Tondl (2004): Regional Convergence in the European Union, 1985-1999: A Spatial Dynamic Panel Analysis. *Regional Studies*, **38**, 241–253.
- Baltagi, B. (2008): *Econometric Analysis of Panel Data*, 4th edition. New York: John Wiley and Sons.
- Bera, A. and M. Yoon, (1993): Specification Testing with Locally Misspecified Alternatives. *Econometric Theory* **9**, 649-658.
- Beenstock, M. and D. Felsenstein (2007): Spatial Vector Autoregressions. *Spatial Economic Analysis* **2**, 167-196.
- Beenstock, M. and D. Felsenstein (2008): Regional Heterogeneity, Conditional Convergence and Regional Inequality. *Regional Studies* **42**, 475-488.
- Breusch, T., and Pagan, A. (1980): The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* **47**, 239–254.

- Burridge, P. (1981): Testing for a Common Factor in a Spatial Autoregression Model. *Environment and Planning A*, **13**, 795-800.
- Campos, J, N. Ericsson and D. Hendry (2005). General-to-Specific Modeling: An Overview and Selected Bibliography. In J. Campos, D. Hendry, and N. Ericsson (Eds.), *General-to-Specific Modeling*. Cheltenham: Edward Elgar Publishing.
- Carlino, G. and R. DeFina (1999): The Differential Regional Effects of Monetary Policy: Evidence from the U.S. States. *Journal of Regional Science* **39**, 339-358.
- Charemza, W. and D. Deadman (1997): *New Directions in Econometric Practice: General to Specific Modelling, Cointegration, and Vector Autoregression*. Cheltenham: Edward Elgar Publishing.
- Chen, X. and T. Conley (2001): A New Semiparametric Spatial Model for Panel Time Series. *Journal of Econometrics* **105**, 59-83.
- Coakley, J., A. Fuentes and R. Smith (2002): A Principal Components Approach to Cross-section Dependence in Panels. Working Paper, Department of Economics, Birbeck College. University of London, London.
- Conley, T. (1999): GMM Estimation with Cross-sectional Dependence. *Journal of Econometrics* **92**, 1-45.
- Conley, T. and F. Molinari (2008): Spatial Correlation Robust Inference with Errors in Location or Distance. *Journal of Econometrics* **40**, 76-96.
- Davidson, J. (2000). *Econometric Theory*. Oxford: Blackwell.
- Di Giacinto, V. (2003): Differential Regional Effects of Monetary Policy: A Geographical SVAR Approach. *International Regional Science Review* **26**, 313-341.
- Di Giacinto, V. (2006): A Generalized Space-Time Model with an Application to Regional Unemployment Analysis in Italy. *International Regional Science Review*, **29**, 159-198.
- Driscoll, J. and A. Kraay (1998): Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *The Review of Economics and Statistics* **80**, 549-560.
- Dufour J. and L. Khalaf (2002): Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions. *Journal of Econometrics* **106**, 143–170.
- Dwivedi, T. and K. Srivastava (1978): Optimality of Least Squares in Seemingly Unrelated Regressions Models. *Journal of Econometrics*, **7**, 391-395.
- Elhorst, P. (2001): Dynamic Models in Space and Time. *Geographical Analysis*, **33**, 244-268.
- Elhorst, P. (2003): Specification and Estimation of Spatial Panel Data Models. *International Regional Science Review*, **26**, 244-268.
- Elhorst, P. (2004): Serial and Spatial Error Dependence in Space-Time Models. In Getis A., J. Mur and H. Zoller (Eds.): *Spatial Econometrics and Spatial Statistics*, pp 176-193. Basingstoke: Palgrave Macmillan.
- Elhorst, P. (2005): Unconditional Maximum Likelihood Estimation of Linear and Log-linear Dynamic Models for Spatial Panels. *Geographical Analysis*, **37**, 85-106.
- Elhorst, P. (2008): Serial and spatial error correlation. *Economics Letters* **100**, 422–424
- Elhorst, P., Blien, U. and K. Wolf (2006): New evidence on the wage curve: A spatial panel approach. *International Regional Science Review* **30**: 173-191
- Egger, P. and M. Pfaffermayr (2004): Distance, Trade and FDI: A Hausman-Taylor SUR Approach. *Journal of Applied Econometrics* **16**, 227-246.
- Fingleton B. (2001): Theoretical Economic Geography and Spatial Econometrics: Dynamic Perspectives. *Journal of Economic Geography*, **1**, 201-225.
- Fingleton B. (2007): Multi-equation spatial econometric model, with application to EU manufacturing productivity growth. *Journal of Geographical Systems* **9**, 119-144.
- Florax R, H, Folmer and S. Rey (2003): Specification Searches in Spatial Econometrics: the Relevance of Hendry's Methodology. *Regional Science and Urban Economics*, **33**, 557-579

- Florax R, H, Folmer and S. Rey (2006): A Comment on “Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology: A Reply. *Regional Science and Urban Economics*, **36**, 300-308.
- Hendry, D. (1980): Econometrics-Alchemy or Science? *Economica*, **47**, 387-406
- Hendry, D. (2006): A Comment on “Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology” *Regional Science and Urban Economics*, **36**, 309-312
- Hordijk, L. (1979): Problems in Estimating Econometric Relations in Space. *Papers, Regional Science Association*, **42**, 99-115.
- Hordijk, L. and P. Nijkamp (1977): Dynamic Models of Spatial Autocorrelation. *Environment and Planning A*, **9**, 505-519.
- Kapoor, M., H. Kelejian and I. Prucha (2007): Panel Data Models with Spatially Correlated Error Components. *Journal of Econometrics*, **140**, 97-130.
- Kosfeld R. and D. Christian (2006): Thresholds for employment and unemployment: A spatial analysis of German regional labour markets, 1992–2000. *Papers in Regional Science*, **85**, 523-542
- Lauridsen, J. Bech, M. López, F. and M. Maté (2010): A spatiotemporal analysis of public pharmaceutical expenditures. *Annals of Regional Science*, **44**, 299-314.
- Lee L. (2004): Asymptotic Distributions of Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica* **72**: 1899-1925
- LeGallo, J. and S. Dall’erba (2006): Evaluating the temporal and spatial heterogeneity of the European convergence process: 1980-1999, *Journal of Regional Science*, **46**, 269-288.
- LeGallo, J. and C. Chasco (2008): Spatial Analysis of Urban Growth in Spain, 1900-2001. *Empirical Economics*, **34**, 59-80.
- Lundberg, J. (2006): Spatial interaction model of spillovers from locally provided public services *Regional Studies*, **40**, 631-644.
- Lütkepohl, H. (2007): General-to-specific or specific-to-general modelling? An Opinion on Current Econometric Terminology. *Journal of Econometrics*, **136**, 319-324
- Malinvaud, E. (1970): *Statistical Methods of Econometrics*. Amsterdam: North Holland.
- Moscone, F., Tosetti, E. and M. Knapp (2007): SUR model with spatial effects: An application to mental health expenditure. *Health Economics*, **16**, 1403-1408.
- Mur, J. and A. Angulo (2006). The Spatial Durbin Model and the Common Factor Tests. *Spatial Economic Analysis* **2**, 207-226.
- Mur, J. and A. Angulo (2009): Model Selection Strategies in a Spatial Setting: Some Additional Results. *Regional Science and Urban Economics* **39** 200-213.
- Mur, J. and F. López (2010): ‘Testing for spatial effects in Seemingly Unrelated Regressions’. Working Paper GAEC/01. (http://metodos.upct.es/detaer/Principal_espanol/wp/index.html)
- Pesaran, H. (2005): Estimation and Inference in Large Heterogeneous Panels with Cross Section Dependence. Technical Report, Department of Economics. University of Cambridge, Cambridge.
- Rey, S. and B. Montouri (1999): US Regional Income Convergence: A Spatial Econometrics Perspective. *Regional Studies*, **33**, 143-156.
- Schmidt, P. (1976): *Econometrics*. New York: Marcel Dekker.
- Theil, H. (1971): *Principles of Econometrics*. New York: John Wiley and Sons.
- Wang, X. and K. Kockelman (2007): Specification and Estimation of a Spatially and Temporally Autocorrelated Seemingly Unrelated Regression Model: Application to Crash Rates in China. *Transportation*, **34**, 281-300.
- White, E. and G. Hewings (1982): Space-Time Employment Modeling: Some Results Using Seemingly Unrelated Regression Estimators. *Journal of Regional Science*, **22**, 283-302.
- Zellner, A. (1962): An Efficient Method of Estimating Seemingly Unrelated Regressions and Test of Aggregation Bias. *Journal of the American Statistical Association*, **57**, 348-68.

Appendix: Results of the ML estimation of the SUR models with spatial effects

This Appendix contains additional results on the ML estimation of the different SUR models introduced in Sections 3, 4 and 5. Specifically, we focus on the expressions of the score vector and of the information matrix for the different models and on the corresponding Lagrange Multipliers.

Section A.I. The SUR-SARAR model.

As indicated, the compact expression of this model is:

$$\left. \begin{aligned} \mathbf{A}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{B}\mathbf{u} &= \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim \mathbf{N}(0, \boldsymbol{\Omega}) \end{aligned} \right\} \quad (A1)$$

$$\mathbf{A} = \mathbf{I}_T \otimes [\mathbf{I}_{GR} - \boldsymbol{\Lambda} \otimes \mathbf{W}]$$

$$\mathbf{B} = \mathbf{I}_T \otimes [\mathbf{I}_{GR} - \boldsymbol{\Upsilon} \otimes \mathbf{W}]$$

$$\boldsymbol{\Omega} = \mathbf{I}_T \otimes \boldsymbol{\Sigma} \otimes \mathbf{I}_R$$

The logarithm of the likelihood function of the model of (A1) is the following:

$$l(\mathbf{y}; \boldsymbol{\theta}) = -\frac{RTG}{2} \ln(2\pi) - \frac{RT}{2} \ln|\boldsymbol{\Sigma}| + T \left[\sum_{g=1}^G \ln|\mathbf{B}_g| + \sum_{g=1}^G \ln|\mathbf{A}_g| \right] - \frac{(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{B}' (\mathbf{I}_T \otimes \boldsymbol{\Sigma} \otimes \mathbf{I}_R)^{-1} \mathbf{B} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2} \quad (A2)$$

where $\boldsymbol{\Omega}^{-1} = \mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R$; $\boldsymbol{\theta}' = [\boldsymbol{\beta}'; \lambda_1; \dots; \lambda_G; \rho_1; \dots; \rho_G; \sigma_{ij}]$ is the vector of parameters of the model, of order $(k+2G+G(G+1)/2) \times 1$. The score vector has the following structure:

$$\mathbf{g}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}} \\ \frac{\partial l}{\partial \lambda_g} \\ \frac{\partial l}{\partial \rho_g} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \mathbf{B}' \boldsymbol{\Omega}^{-1} \mathbf{B} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ -T \text{tr} [\mathbf{A}_g^{-1} \mathbf{W}] + (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{B}' \boldsymbol{\Omega}^{-1} \mathbf{B} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) \mathbf{y} \\ -T \text{tr} [\mathbf{B}_g^{-1} \mathbf{W}] + (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{B}' \boldsymbol{\Omega}^{-1} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ -\frac{RT}{2} \text{tr} [\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij}] + \frac{(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{B}' [\mathbf{I}_T \otimes (\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1}) \otimes \mathbf{I}_R] \mathbf{B} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2} \end{bmatrix} \quad (A3)$$

where \mathbf{E}^{ij} (analogously \mathbf{E}^{gg}) is a $(G \times G)$ matrix whose elements are all zero except the (i,j) and the (j,i) which are 1. The set of second derivatives is:

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\mathbf{X}' \mathbf{B}' \boldsymbol{\Omega}^{-1} \mathbf{B} \mathbf{X}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \lambda_g} = -\mathbf{X}' \mathbf{B}' \boldsymbol{\Omega}^{-1} \mathbf{B} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) \mathbf{y}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \rho_g} = -\mathbf{X}' \left[\mathbf{B}' \left\{ \mathbf{I}_T \otimes (\boldsymbol{\Sigma}^{-1} \mathbf{E}^{gg}) \otimes \mathbf{W} \right\} + \left\{ \mathbf{I}_T \otimes (\mathbf{E}^{gg} \boldsymbol{\Sigma}^{-1}) \otimes \mathbf{W} \right\} \mathbf{B} \right] (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \sigma_{ij}} = -\mathbf{X}' \mathbf{B}' \left[\mathbf{I}_T \otimes (\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1}) \otimes \mathbf{I}_R \right] \mathbf{B} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$
(A4.a)

$$\begin{aligned}
\frac{\partial^2 1}{\partial \lambda_g^2} &= -\text{Ttr}[\mathbf{A}_g^{-1} \mathbf{W} \mathbf{A}_g^{-1} \mathbf{W}] - y' (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W})' \mathbf{B}' \Omega^{-1} \mathbf{B} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) y \\
\frac{\partial^2 1}{\partial \lambda_g \partial \lambda_s} &= -y' (\mathbf{I}_T \otimes \mathbf{E}^{ss} \otimes \mathbf{W})' \mathbf{B}' \Omega^{-1} \mathbf{B} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) y \\
\frac{\partial^2 1}{\partial \lambda_g \partial \rho_g} &= -(\mathbf{A}y - \mathbf{X}\beta)' \left[(\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W})' \Omega^{-1} \mathbf{B} + \mathbf{B}' \Omega^{-1} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) \right] (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) y \\
\frac{\partial^2 1}{\partial \lambda_g \partial \rho_s} &= -(\mathbf{A}y - \mathbf{X}\beta)' \left[(\mathbf{I}_T \otimes \mathbf{E}^{ss} \otimes \mathbf{W})' \Omega^{-1} \mathbf{B} + \mathbf{B}' \Omega^{-1} (\mathbf{I}_T \otimes \mathbf{E}^{ss} \otimes \mathbf{W}) \right] (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) y \\
\frac{\partial^2 1}{\partial \lambda_g \partial \sigma_{ij}} &= -(\mathbf{A}y - \mathbf{X}\beta)' \mathbf{B}' \left[\mathbf{I}_T \otimes (\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1}) \otimes \mathbf{I}_R \right] \mathbf{B} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) y
\end{aligned} \tag{A4.b}$$

$$\begin{aligned}
\frac{\partial^2 1}{\partial \rho_g^2} &= -\text{Ttr}[\mathbf{B}_g^{-1} \mathbf{W} \mathbf{B}_g^{-1} \mathbf{W}] - (\mathbf{A}y - \mathbf{X}\beta)' (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W})' \Omega^{-1} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) (\mathbf{A}y - \mathbf{X}\beta) \\
\frac{\partial^2 1}{\partial \rho_g \partial \rho_s} &= -(\mathbf{A}y - \mathbf{X}\beta)' (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W})' \Omega^{-1} (\mathbf{I}_T \otimes \mathbf{E}^{ss} \otimes \mathbf{W}) (\mathbf{A}y - \mathbf{X}\beta)
\end{aligned} \tag{A4.c}$$

$$\begin{aligned}
\frac{\partial^2 1}{\partial \rho_g \partial \sigma_{ij}} &= -(\mathbf{A}y - \mathbf{X}\beta)' (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W})' \left[\mathbf{I}_T \otimes (\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1}) \otimes \mathbf{I}_R \right] \mathbf{B} (\mathbf{A}y - \mathbf{X}\beta) \\
\frac{\partial^2 1}{\partial \sigma_{ij}^2} &= \frac{\text{RT}}{2} \text{tr}[\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij}] - (\mathbf{A}y - \mathbf{X}\beta)' \mathbf{B}' \left[\mathbf{I}_T \otimes (\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1}) \otimes \mathbf{I}_R \right] \mathbf{B} (\mathbf{A}y - \mathbf{X}\beta) \\
\frac{\partial^2 1}{\partial \sigma_{ij} \partial \sigma_{sr}} &= \frac{\text{RT}}{2} \text{tr}[\boldsymbol{\Sigma}^{-1} \mathbf{E}^{sr} \boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij}] - (\mathbf{A}y - \mathbf{X}\beta)' \mathbf{B}' \left[\mathbf{I}_T \otimes (\boldsymbol{\Sigma}^{-1} \mathbf{E}^{sr} \boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1}) \otimes \mathbf{I}_R \right] \mathbf{B} (\mathbf{A}y - \mathbf{X}\beta)
\end{aligned} \tag{A4.d}$$

The expected value of these terms, changing the sign, is⁸:

$$\begin{aligned}
\mathbf{I}_{\beta\beta} &= \mathbf{X}' \mathbf{B}' \Omega^{-1} \mathbf{B} \mathbf{X} \\
\mathbf{I}_{\beta\lambda_g} &= \mathbf{X}' \mathbf{B}' \Omega^{-1} \mathbf{B} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) \mathbf{A}^{-1} \mathbf{X} \beta; \quad g = 1, 2, \dots, G \\
\mathbf{I}_{\beta\rho_g} &= 0; \quad g = 1, 2, \dots, G \\
\mathbf{I}_{\beta\sigma_{ij}} &= 0; \quad i, j = 1, 2, \dots, G
\end{aligned} \tag{A5.a}$$

⁸ For brevity's sake, we use the notation $\mathbf{I}_{\eta\gamma} = -\mathbf{E} \left[\frac{\partial^2 1}{\partial \eta \partial \gamma'} \right]$.

$$I_{\lambda_g \lambda_g} = \text{Tr} \left[\mathbf{A}_g^{-1} \mathbf{W} \mathbf{A}_g^{-1} \mathbf{W}' \right] + \beta' \mathbf{X}' \mathbf{H}_{\text{ARAR}}^{\text{gg}} \mathbf{X} \beta + \text{tr} \mathbf{H}_{\text{ARAR}}^{\text{gg}} \mathbf{B}^{-1} \Omega \mathbf{B}^{-1}; \quad g = 1, 2, \dots, G$$

$$I_{\lambda_g \lambda_s} = \beta' \mathbf{X}' \mathbf{H}_{\text{ARAR}}^{\text{gs}} \mathbf{X} \beta + \text{tr} \mathbf{H}_{\text{ARAR}}^{\text{gs}} \mathbf{B}^{-1} \Omega \mathbf{B}^{-1}; \quad g, s = 1, 2, \dots, G$$

$$I_{\lambda_g \rho_s} = \text{tr} \left\{ \left[\Omega \mathbf{B}^{-1} \left(\mathbf{I}_T \otimes \mathbf{E}^{\text{gg}} \Sigma^{-1} \otimes \mathbf{W}' \right) \mathbf{B} \left(\mathbf{I}_T \otimes \mathbf{E}^{\text{ss}} \otimes \mathbf{W} \right) + \left(\mathbf{I}_T \otimes \mathbf{E}^{\text{gg}} \mathbf{E}^{\text{ss}} \otimes \mathbf{W} \mathbf{W}' \right) \right] (\mathbf{B} \mathbf{A})^{-1} \right\}; \quad g, s = 1, 2, \dots, G$$

$$I_{\lambda_g \sigma_{ij}} = \text{tr} \left[\mathbf{I}_T \otimes \left(\mathbf{E}^{ij} \Sigma^{-1} \right) \otimes \mathbf{I}_R \right] \mathbf{B} \left(\mathbf{I}_T \otimes \mathbf{E}^{\text{gg}} \otimes \mathbf{W} \right) (\mathbf{B} \mathbf{A})^{-1}; \quad g; i, j = 1, 2, \dots, G$$

(A5.b)

$$I_{\rho_g \rho_g} = \text{Tr} \left[\mathbf{B}_g^{-1} \mathbf{W} \mathbf{B}_g^{-1} \mathbf{W}' \right] + \text{tr} \mathbf{B}^{-1} \left[\mathbf{I}_T \otimes \left(\mathbf{E}^{\text{gg}} \Sigma^{-1} \mathbf{E}^{\text{gg}} \right) \otimes \left(\mathbf{W}' \mathbf{W} \right) \right] \mathbf{B}^{-1} \Omega; \quad g = 1, 2, \dots, G$$

$$I_{\rho_g \rho_s} = \text{tr} \mathbf{B}^{-1} \left[\mathbf{I}_T \otimes \left(\mathbf{E}^{\text{ss}} \Sigma^{-1} \mathbf{E}^{\text{gg}} \right) \otimes \left(\mathbf{W}' \mathbf{W} \right) \right] \mathbf{B}^{-1} \Omega; \quad g, s = 1, 2, \dots, G$$

(A5.c)

$$I_{\rho_g \sigma_{ij}} = \text{tr} \left[\mathbf{I}_T \otimes \left(\mathbf{E}^{\text{gg}} \Sigma^{-1} \mathbf{E}^{ij} \right) \otimes \mathbf{W} \right] \mathbf{B}^{-1}; \quad g; i, j = 1, 2, \dots, G$$

$$I_{\sigma_{ij} \sigma_{sr}} = \frac{\text{TR}}{2} \text{tr} \left[\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1} \mathbf{E}^{sr} \right]; \quad i, j, s, r = 1, 2, \dots, G$$

(A5.d)

where $\mathbf{H}_{\text{ARAR}}^{\text{gs}} = \mathbf{A}^{-1} \left(\mathbf{I}_T \otimes \mathbf{E}^{\text{ss}} \otimes \mathbf{W} \right)' \mathbf{B}' \left[\mathbf{I}_T \otimes \Sigma^{-1} \otimes \mathbf{I}_R \right] \mathbf{B} \left(\mathbf{I}_T \otimes \mathbf{E}^{\text{gg}} \otimes \mathbf{W} \right) \mathbf{A}^{-1}$ and σ^{st} is the element (t,s) of the matrix Σ^{-1} . We introduce the following ordering of the information matrix:

$$I(\theta) = \begin{bmatrix} I_{\beta\beta} & I_{\beta\lambda} & I_{\beta\rho} & I_{\beta\sigma} \\ (k \times k) & (k \times T) & (k \times T) & (k \times (T(T+1)/2)) \\ & I_{\lambda\lambda} & I_{\lambda\rho} & I_{\lambda\sigma} \\ & (T \times T) & (T \times T) & (T \times (T(T+1)/2)) \\ & & I_{\rho\rho} & I_{\rho\sigma} \\ & & (T \times T) & (T \times (T(T+1)/2)) \\ & & & I_{\sigma\sigma} \\ & & & ((T(T+1)/2) \\ & & & x(T(T+1)/2)) \end{bmatrix} \quad (\text{A6})$$

The null hypothesis that there are no spatial effects in the SUR model of (A1) is:

$$\left. \begin{array}{l} H_0 : \lambda_g = \rho_g = 0; \forall g \\ H_A : \text{No } H_0 \end{array} \right\}$$

$$\text{Under } H_0 \Rightarrow \mathbf{A} = \mathbf{B} = \mathbf{I}_{\text{GTR}}$$

(A7)

$$\left. \begin{array}{l} y = \mathbf{X}\beta + \mathbf{u} \\ \Rightarrow \mathbf{u} = \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim N(0, \Omega) \Rightarrow \Omega = \mathbf{I}_T \otimes \Sigma \otimes \mathbf{I}_R \end{array} \right\}$$

The score of (A3) becomes:

$$\mathbf{g}(\theta)_{|H_0} = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \lambda_g} \\ \frac{\partial l}{\partial \rho_g} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} \mathbf{g}(\beta)_{|H_0} \\ \mathbf{g}(\lambda)_{|H_0} \\ \mathbf{g}(\rho)_{|H_0} \\ \mathbf{g}(\sigma)_{|H_0} \end{bmatrix} = \begin{bmatrix} 0 \\ \hat{\mathbf{u}}' \left[\mathbf{I}_T \otimes \left(\boldsymbol{\Sigma}^{-1} \mathbf{E}^{gg} \right) \otimes \mathbf{W} \right] \mathbf{y} \\ \hat{\mathbf{u}}' \left[\mathbf{I}_T \otimes \left(\boldsymbol{\Sigma}^{-1} \mathbf{E}^{gg} \right) \otimes \mathbf{W} \right] \hat{\mathbf{u}} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \sum_{t=1}^T \sum_{s=1}^G \sigma^{sg} \hat{\mathbf{u}}'_{tg} \mathbf{W} \mathbf{y}_{ts} \\ \sum_{t=1}^T \sum_{s=1}^G \sigma^{sg} \hat{\mathbf{u}}'_{tg} \mathbf{W} \hat{\mathbf{u}}_{ts} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \sum_{s=1}^G \sigma^{sg} \text{tr}(\hat{\mathbf{u}}'_g \mathbf{y}_{Ls}) \\ \sum_{s=1}^G \sigma^{sg} \text{tr}(\hat{\mathbf{u}}'_g \hat{\mathbf{u}}_{Ls}) \\ 0 \end{bmatrix} \quad (\text{A8})$$

$\hat{\mathbf{u}}$ is the (TGRx1) vector of residuals of the SUR model, estimated in the absence of spatial effects. The score is made up of two sub-vectors of zeros, of orders (kx1) and (G(G+1)/2)x1, respectively, and another two non-zero sub-vectors, both of order (Gx1), as appears in (A8). Furthermore

$$\begin{aligned} \hat{\mathbf{u}}_g &= \left[\hat{\mathbf{u}}_{1g} \quad \hat{\mathbf{u}}_{2g} \quad \cdots \quad \hat{\mathbf{u}}_{Tg} \right]_{(\text{RxTR})} & \mathbf{y}_L &= \left[\mathbf{y}_{L1} \quad \mathbf{y}_{L2} \quad \cdots \quad \mathbf{y}_{LG} \right]_{(\text{RxTRG})} \\ \mathbf{y}_{Lg} &= \left[\mathbf{W} \mathbf{y}_{g1} \quad \mathbf{W} \mathbf{y}_{g2} \quad \cdots \quad \mathbf{W} \mathbf{y}_{gT} \right]_{(\text{RxTR})} & \hat{\mathbf{u}}_L &= \left[\hat{\mathbf{u}}_{L1} \quad \hat{\mathbf{u}}_{L2} \quad \cdots \quad \hat{\mathbf{u}}_{LG} \right]_{(\text{RxTRG})} \\ \hat{\mathbf{u}}_{Lg} &= \left[\mathbf{W} \hat{\mathbf{u}}_{g1} \quad \mathbf{W} \hat{\mathbf{u}}_{g2} \quad \cdots \quad \mathbf{W} \hat{\mathbf{u}}_{gT} \right]_{(\text{RxTR})} \end{aligned}$$

The elements of the information matrix, also under the null hypothesis, are:

$$\begin{aligned} I_{\beta\beta} &= \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X} \\ I_{\beta\lambda_g} &= \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \mathbf{E}^{gg} \otimes \mathbf{W} \right] \mathbf{X} \beta; \quad g = 1, 2, \dots, G \\ I_{\beta\rho_g} &= 0; \quad t = 1, 2, \dots, G \\ I_{\beta\sigma_{ij}} &= 0; \quad i, j = 1, 2, \dots, G \end{aligned} \quad (\text{A9.a})$$

$$\begin{aligned} I_{\lambda_g \lambda_g} &= \sigma^{gg} \left(\beta' \mathbf{X}' \left[\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes (\mathbf{W}' \mathbf{W}) \right] \mathbf{X} \beta + \sigma_{gg} \text{tr}(\mathbf{W}' \mathbf{W}) \right) + T \text{tr} \mathbf{W}^2; \quad g = 1, 2, \dots, G \\ I_{\lambda_g \lambda_s} &= \sigma^{gs} \left(\beta' \mathbf{X}' \left[\mathbf{I}_T \otimes \mathbf{E}^{gs} \otimes (\mathbf{W}' \mathbf{W}) \right] \mathbf{X} \beta + \sigma_{gs} \text{tr}(\mathbf{W}' \mathbf{W}) \right); \quad g, s = 1, 2, \dots, G \\ I_{\lambda_g \rho_g} &= T \sigma^{gg} \sigma_{gg} \left[\text{tr}(\mathbf{W}' \mathbf{W}) + \text{tr}(\mathbf{W} \mathbf{W}) \right]; \quad g = 1, 2, \dots, G \\ I_{\lambda_g \rho_s} &= T \sigma^{gs} \sigma_{gs} \left[\text{tr}(\mathbf{W}' \mathbf{W}) + \text{tr}(\mathbf{W} \mathbf{W}) \right]; \quad g, s = 1, 2, \dots, G \end{aligned} \quad (\text{A9.b})$$

$$\begin{aligned} I_{\lambda_g \sigma_{ij}} &= 0; \quad g; i, j = 1, 2, \dots, G \\ I_{\rho_g \rho_g} &= T \left[\text{tr}(\mathbf{W}' \mathbf{W}) \sigma^{gg} \sigma_{gg} + \text{tr}(\mathbf{W} \mathbf{W}) \right]; \quad g = 1, 2, \dots, G \\ I_{\rho_g \rho_s} &= T \sigma^{gs} \sigma_{gs} \text{tr}(\mathbf{W}' \mathbf{W}); \quad g, s = 1, 2, \dots, G \\ I_{\rho_g \sigma_{ij}} &= 0; \quad g; i, j = 1, 2, \dots, G \end{aligned} \quad (\text{A9.c})$$

$$I_{\sigma_{ij}\sigma_{sr}} = \frac{\text{TR}}{2} \text{tr} \left[\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1} \mathbf{E}^{sr} \right]; \quad i, j, s, r = 1, 2, \dots, G \quad (\text{A9.d})$$

All these elements have been defined previously. To sum up, the information matrix becomes:

$$I(\theta)_{|H_0} = \begin{bmatrix} I_{\beta\beta} & I_{\beta\lambda} & 0 & 0 \\ (k \times k) & (k \times G) & (k \times G) & (k \times (G(G+1)/2)) \\ & I_{\lambda\lambda} & I_{\lambda\rho} & 0 \\ & (G \times G) & (G \times G) & (G \times (G(G+1)/2)) \\ & & I_{\rho\rho} & 0 \\ & & (G \times G) & (G \times (G(G+1)/2)) \\ & & & I_{\sigma\sigma} \\ & & & ((G(G+1)/2) \\ & & & x(G(G+1)/2)) \end{bmatrix} \quad (\text{A10})$$

This matrix is block-diagonal:

$$I(\theta)_{|H_0} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \Rightarrow \begin{cases} M_{11} = \begin{bmatrix} I_{\beta\beta} & I_{\beta\lambda} & 0 \\ (k \times k) & (k \times G) & (k \times G) \\ & I_{\lambda\lambda} & I_{\lambda\rho} \\ & (G \times G) & (G \times G) \\ & & I_{\rho\rho} \\ & & (G \times G) \end{bmatrix} \\ M_{22} = \begin{bmatrix} I_{\sigma\sigma} \\ ((G(G-1)/2) \\ x(G(G-1)/2)) \end{bmatrix} \\ M_{12} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ M_{21} = [0 \quad 0 \quad 0] \end{cases} \quad (\text{A11})$$

The sub-matrix $I_{\lambda\rho}$ is diagonal, which implies that the ML estimators of ρ_g and of λ_s , under the null hypothesis, are correlated for the same equation but they are independent for different equations; that is, $\text{Cov}(\rho_g, \lambda_s) = 0$ if $g \neq s$ and $\text{Cov}(\rho_g, \lambda_s) \neq 0$ if $g=s$. The Lagrange Multiplier, for the hypothesis of (A7), is the quadratic form of the score evaluated in the null hypothesis (as in A8), on the inverse of the information matrix, also evaluated in the null hypothesis (as in A9 and A10). The final result is:

$$\begin{aligned} \mathbf{LM}_{\text{SARAR}}^{\text{SUR}} &= \left[\mathbf{g}(\theta)_{|H_0} \right]' \left[I(\theta)_{|H_0} \right]^{-1} \left[\mathbf{g}(\theta)_{|H_0} \right]_{\text{as}} \sim \chi^2(2G) \\ \Rightarrow \mathbf{LM}_{\text{SARAR}}^{\text{SUR}} &= \left[\mathbf{g}'(\lambda)_{|H_0} \quad \mathbf{g}'(\rho)_{|H_0} \right] \begin{bmatrix} I_{\lambda\lambda} - I_{\lambda\beta} I_{\beta\beta}^{-1} I_{\beta\lambda} & I_{\lambda\rho} \\ I_{\rho\lambda} & I_{\rho\rho} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}(\lambda)_{|H_0} \\ \mathbf{g}(\rho)_{|H_0} \end{bmatrix}_{\text{as}} \sim \chi^2(2G) \end{aligned} \quad (\text{A12})$$

Section A.II. The SUR-SLM model.

The model that corresponds to this case is:

$$\left. \begin{aligned} \mathbf{A}y &= \mathbf{X}\beta + \varepsilon \\ \varepsilon &\sim N(0, \Omega) \Rightarrow \Omega = \mathbf{I}_T \otimes \Sigma \otimes \mathbf{I}_R \end{aligned} \right\} \quad (\text{A13})$$

$$\mathbf{A} = \mathbf{I}_T \otimes [\mathbf{I}_{GR} - \Lambda \otimes \mathbf{W}]$$

Matrix Ω appears in (A1). The logarithm of the likelihood function, introducing the SLM structure of (A13), is:

$$l(y; \theta) = -\frac{RTG}{2} \ln(2\pi) - \frac{RT}{2} \ln|\Sigma| + T \sum_{g=1}^G \ln|\mathbf{A}_g| - \frac{(\mathbf{A}y - \mathbf{X}\beta)' (\mathbf{I}_T \otimes \Sigma \otimes \mathbf{I}_R)^{-1} (\mathbf{A}y - \mathbf{X}\beta)}{2} \quad (\text{A14})$$

where $\theta' = [\beta'; \lambda_1; \dots; \lambda_G; \sigma_{ij}]$ is the vector of parameters of the model, of order $(k+G+G(G+1)/2) \times 1$.

The score vector has the following structure:

$$\mathbf{g}(\theta) = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \lambda_g} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \Omega^{-1} (\mathbf{A}y - \mathbf{X}\beta) \\ -T \text{tr} [\mathbf{A}_t^{-1} \mathbf{W}] + (\mathbf{A}y - \mathbf{X}\beta)' \Omega^{-1} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) y \\ -\frac{\text{TR}}{2} \text{tr} [\Sigma^{-1} \mathbf{E}^{ij}] + \frac{(\mathbf{A}y - \mathbf{X}\beta)' [\mathbf{I}_T \otimes (\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1}) \otimes \mathbf{I}_R] (\mathbf{A}y - \mathbf{X}\beta)}{2} \end{bmatrix} \quad (\text{A15})$$

The second derivatives corresponding to this case are:

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta \partial \beta'} &= -\mathbf{X}' \Omega^{-1} \mathbf{X} \\ \frac{\partial^2 l}{\partial \beta \partial \lambda_g} &= -\mathbf{X}' \Omega^{-1} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) y \\ \frac{\partial^2 l}{\partial \beta \partial \sigma_{ij}} &= -\mathbf{X}' [\mathbf{I}_T \otimes (\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1}) \otimes \mathbf{I}_R] (\mathbf{A}y - \mathbf{X}\beta) \end{aligned} \quad (\text{A16.a})$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \lambda_g^2} &= -T \text{tr} [\mathbf{A}_g^{-1} \mathbf{W} \mathbf{A}_g^{-1} \mathbf{W}] - y' (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W})' \Omega^{-1} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) y \\ \frac{\partial^2 l}{\partial \lambda_g \partial \lambda_s} &= -y' (\mathbf{I}_T \otimes \mathbf{E}^{ss} \otimes \mathbf{W})' \Omega^{-1} (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) y \\ \frac{\partial^2 l}{\partial \lambda_g \partial \sigma_{ij}} &= -(\mathbf{A}y - \mathbf{X}\beta)' [\mathbf{I}_T \otimes (\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1}) \otimes \mathbf{I}_R] (\mathbf{I}_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) y \end{aligned} \quad (\text{A16.b})$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \sigma_{ij}^2} &= \frac{\text{TR}}{2} \text{tr} [\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1} \mathbf{E}^{ij}] - (\mathbf{A}y - \mathbf{X}\beta)' [\mathbf{I}_T \otimes (\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1}) \otimes \mathbf{I}_R] (\mathbf{A}y - \mathbf{X}\beta) \\ \frac{\partial^2 l}{\partial \sigma_{ij} \partial \sigma_{sr}} &= \frac{\text{TR}}{2} \text{tr} [\Sigma^{-1} \mathbf{E}^{sr} \Sigma^{-1} \mathbf{E}^{ij}] - (\mathbf{A}y - \mathbf{X}\beta)' [\mathbf{I}_T \otimes (\Sigma^{-1} \mathbf{E}^{sr} \Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1}) \otimes \mathbf{I}_R] (\mathbf{A}y - \mathbf{X}\beta) \end{aligned} \quad (\text{A16.c})$$

The expected values, changing the sign, are:

$$I_{\beta\beta} = X' \Omega^{-1} X$$

$$I_{\beta\lambda_g} = X' \Omega^{-1} (I_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) \mathbf{A}^{-1} X \beta; \quad g = 1, 2, \dots, G \quad (\text{A17.a})$$

$$I_{\beta\sigma_{ij}} = 0; \quad i, j = 1, 2, \dots, G$$

$$I_{\lambda_g \lambda_g} = \text{Tr} \left[\mathbf{A}_g^{-1} \mathbf{W} \mathbf{A}_g^{-1} \mathbf{W} \right] + \beta' X' \mathbf{H}_{\text{SLM}}^{gg} X \beta + \text{tr} \mathbf{H}_{\text{SLM}}^{gg} \Omega; \quad g = 1, 2, \dots, G$$

$$I_{\lambda_g \lambda_s} = \beta' X' \mathbf{H}_{\text{SLM}}^{gs} X \beta + \text{tr} \mathbf{H}_{\text{SLM}}^{gs} \Omega; \quad g, s = 1, 2, \dots, G \quad (\text{A17.b})$$

$$I_{\lambda_g \sigma_{ij}} = \text{tr} \left[I_T \otimes (\mathbf{E}^{ij} \Sigma^{-1}) \otimes I_R \right] (I_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) \mathbf{A}^{-1}; \quad g; i, j = 1, 2, \dots, G$$

$$I_{\sigma_{ij} \sigma_{sr}} = \frac{\text{TR}}{2} \text{tr} \left[\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1} \mathbf{E}^{sr} \right]; \quad i, j, s, r = 1, 2, \dots, G \quad (\text{A17.c})$$

where $\mathbf{H}_{\text{SLM}}^{gs} = \mathbf{A}^{-1} (I_T \otimes \mathbf{E}^{ss} \otimes \mathbf{W})' [I_T \otimes \Sigma^{-1} \otimes I_R] (I_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) \mathbf{A}^{-1}$ and σ^{sg} , as before, is the element (g,s) of matrix Σ^{-1} . Now we use the following ordering of the information matrix:

$$I(\theta) = \begin{bmatrix} I_{\beta\beta} & I_{\beta\lambda} & I_{\beta\sigma} \\ (k \times k) & (k \times G) & (k \times (G(G+1)/2)) \\ & I_{\lambda\lambda} & I_{\lambda\sigma} \\ & (G \times G) & (G \times (G(G+1)/2)) \\ & & I_{\sigma\sigma} \\ & & ((G(G+1)/2) \\ & & \times (G(G+1)/2)) \end{bmatrix} \quad (\text{A.18})$$

The null hypothesis in which we are interested is that the spatial lag of the endogenous variable that appears in the right-hand side of the SUR model of (A13) is not relevant:

$$\left. \begin{array}{l} H_0: \lambda_g = 0; \forall g \\ H_A: \text{No } H_0 \end{array} \right\} \quad (\text{A19})$$

$$\text{Under } H_0 \Rightarrow \mathbf{A} = I_{\text{GTR}} \Rightarrow \left. \begin{array}{l} y = X\beta + \varepsilon \\ \varepsilon \sim N(0, \Omega) \Rightarrow \Omega = I_T \otimes \Sigma \otimes I_R \end{array} \right\}$$

The score of (A15) becomes:

$$g(\theta)_{H_0} = \begin{bmatrix} \frac{\partial}{\partial \beta} \\ \frac{\partial}{\partial \lambda_g} \\ \frac{\partial}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} g(\beta)_{H_0} \\ g(\lambda)_{H_0} \\ g(\sigma)_{H_0} \end{bmatrix} = \begin{bmatrix} 0 \\ \hat{u}' [I_T \otimes (\Sigma^{-1} \mathbf{E}^{gg}) \otimes \mathbf{W}] y \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \sum_{t=1}^T \sum_{s=1}^G \sigma^{sg} \hat{u}'_{tg} \mathbf{W} y_{ts} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \sum_{s=1}^G \sigma^{sg} \text{tr}(\hat{u}'_g y_{Ls}) \\ 0 \end{bmatrix} \quad (\text{A20})$$

As before, \hat{u} is the (TGRx1) vector of residuals of the SUR model in the absence of spatial effects and y_L the spatial lag of vector y . The other terms have been defined before. The elements of the information matrix, also under the null hypothesis, are:

$$I_{\beta\beta} = X' \Omega^{-1} X$$

$$I_{\beta\lambda_g} = X' \left[I_T \otimes \Sigma^{-1} E^{gg} \otimes W \right] X \beta; \quad g = 1, 2, \dots, G \quad (A21.a)$$

$$I_{\beta\sigma_{ij}} = 0; \quad i, j = 1, 2, \dots, G$$

$$I_{\lambda_g \lambda_g} = \sigma^{gg} \left[\beta' X' \left(I_T \otimes E^{gg} \otimes (W' W) \right) X \beta + \sigma_{gg} \text{tr}(W' W) \right] + \text{Tr} W^2; \quad g = 1, 2, \dots, G$$

$$I_{\lambda_g \lambda_s} = \sigma^{gs} \left[\beta' X' \left(I_T \otimes E^{gs} \otimes (W' W) \right) X \beta + \sigma_{gs} \text{tr}(W' W) \right]; \quad g, s = 1, 2, \dots, G \quad (A21.b)$$

$$I_{\lambda_g \sigma_{ij}} = 0; \quad g; i, j = 1, 2, \dots, G$$

$$I_{\sigma_{ij} \sigma_{sr}} = \frac{\text{TR}}{2} \text{tr} \left[\Sigma^{-1} E^{ij} \Sigma^{-1} E^{sr} \right]; \quad i, j, s, r = 1, 2, \dots, G \quad (A21.c)$$

This matrix has a block-diagonal structure:

$$I(\theta)_{|H_0} = \begin{bmatrix} I_{\beta\beta} & I_{\beta\lambda} & 0 \\ (k \times k) & (k \times G) & (k \times G) \\ I_{\lambda\lambda} & 0 & 0 \\ (G \times G) & (G \times (G-1)/2) & (G \times (G-1)/2) \\ I_{\sigma\sigma} & & \\ ((G-1)/2) & & \\ x(G-1)/2) & & \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{cases} M_{11} = \begin{bmatrix} I_{\beta\beta} & I_{\beta\lambda} \\ (k \times k) & (k \times G) \\ & I_{\lambda\lambda} \\ & (G \times G) \end{bmatrix} \\ M_{22} = \begin{bmatrix} I_{\sigma\sigma} \\ ((G-1)/2) \\ x(G-1)/2) \end{bmatrix} \\ M_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ M_{21} = \begin{bmatrix} 0 & 0 \end{bmatrix} \end{cases} \quad (A22)$$

Finally, the Lagrange Multiplier, emerges as:

$$\mathbf{LM}_{\text{SLM}}^{\text{SUR}} = \left[\mathbf{g}(\theta)_{|H_0} \right]' \left[I(\theta)_{|H_0} \right]^{-1} \left[\mathbf{g}(\theta)_{|H_0} \right]_{\text{as}} \sim \chi^2(G) \quad (A23)$$

$$\Rightarrow \mathbf{LM}_{\text{SLM}}^{\text{SUR}} = \mathbf{g}'_{(\lambda)_{|H_0}} \left[I_{\lambda\lambda} - I_{\lambda\beta} I_{\beta\beta}^{-1} I_{\beta\lambda} \right]^{-1} \mathbf{g}_{(\lambda)_{|H_0}} \sim \chi^2(G)$$

Section A.III The SUR-SEM model.

The specification corresponding to this model is:

$$\left. \begin{aligned} y &= X\beta + u \\ \mathbf{B}u &= \varepsilon \\ \varepsilon &\sim N(0, \Omega) \Rightarrow \Omega = I_T \otimes \Sigma \otimes I_R \end{aligned} \right\} \quad (A24)$$

$$\mathbf{B} = I_T \otimes [I_{GR} - \Upsilon \otimes \mathbf{W}]$$

The logarithm of the likelihood function is:

$$l(y; \theta) = -\frac{RTG}{2} \ln(2\pi) - \frac{TR}{2} \ln|\Sigma| + T \sum_{g=1}^G \ln|\mathbf{B}_g| - \frac{(y - X\beta)' \mathbf{B}' (I_T \otimes \Sigma \otimes I_R)^{-1} \mathbf{B} (y - X\beta)}{2} \quad (\text{A25})$$

where $\theta' = [\beta'; \rho_1; \dots; \rho_G; \sigma_{ij}]$ is a vector of parameters, of order $(k+G+G(G+1)/2) \times 1$. The score vector has the following composition:

$$\mathbf{g}(\theta) = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \rho_t} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \mathbf{B}' \Omega^{-1} \mathbf{B} (y - X\beta) \\ -\text{Tr}[\mathbf{B}_g^{-1} \mathbf{W}] + (y - X\beta)' \mathbf{B}' \Omega^{-1} (I_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) (y - X\beta) \\ -\frac{TR}{2} \text{tr}[\Sigma^{-1} \mathbf{E}^{ij}] + \frac{(y - X\beta)' \mathbf{B}' [I_T \otimes (\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1}) \otimes I_R] \mathbf{B} (y - X\beta)}{2} \end{bmatrix} \quad (\text{A26})$$

The second derivatives of the log-likelihood function are:

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta \partial \beta'} &= -\mathbf{X}' \mathbf{B}' \Omega^{-1} \mathbf{B} \mathbf{X} \\ \frac{\partial^2 l}{\partial \beta \partial \rho_g} &= -\mathbf{X}' \left[\mathbf{B}' \left\{ I_T \otimes (\Sigma^{-1} \mathbf{E}^{gg}) \otimes \mathbf{W} \right\} + \left\{ I_T \otimes (\mathbf{E}^{gg} \Sigma^{-1}) \otimes \mathbf{W}' \right\} \mathbf{B} \right] (y - X\beta) \end{aligned} \quad (\text{A27.a})$$

$$\frac{\partial^2 l}{\partial \beta \partial \sigma_{ij}} = -\mathbf{X}' \mathbf{B}' \left[I_T \otimes (\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1}) \otimes I_R \right] \mathbf{B} (y - X\beta)$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \rho_g^2} &= -\text{Tr}[\mathbf{B}_g^{-1} \mathbf{W} \mathbf{B}_g^{-1} \mathbf{W}] - (y - X\beta)' (I_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W})' \Omega^{-1} (I_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}) (y - X\beta) \\ \frac{\partial^2 l}{\partial \rho_g \partial \rho_s} &= -(y - X\beta)' (I_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}') \Omega^{-1} (I_T \otimes \mathbf{E}^{ss} \otimes \mathbf{W}) (y - X\beta) \end{aligned} \quad (\text{A27.b})$$

$$\frac{\partial^2 l}{\partial \rho_g \partial \sigma_{ij}} = -(y - X\beta)' (I_T \otimes \mathbf{E}^{gg} \otimes \mathbf{W}') \left[I_T \otimes (\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1}) \otimes I_R \right] \mathbf{B} (y - X\beta)$$

$$\frac{\partial^2 l}{\partial \sigma_{ij}^2} = \frac{TR}{2} \text{tr}[\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1} \mathbf{E}^{ij}] - (y - X\beta)' \mathbf{B}' \left[I_T \otimes (\Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1}) \otimes I_R \right] \mathbf{B} (y - X\beta) \quad (\text{A27.c})$$

$$\frac{\partial^2 l}{\partial \sigma_{ij} \partial \sigma_{sr}} = \frac{TR}{2} \text{tr}[\Sigma^{-1} \mathbf{E}^{sr} \Sigma^{-1} \mathbf{E}^{ij}] - (y - X\beta)' \mathbf{B}' \left[I_T \otimes (\Sigma^{-1} \mathbf{E}^{sr} \Sigma^{-1} \mathbf{E}^{ij} \Sigma^{-1}) \otimes I_R \right] \mathbf{B} (y - X\beta)$$

Their expected value, after changing the sign are:

$$I_{\beta\beta} = \mathbf{X}' \mathbf{B}' \Omega^{-1} \mathbf{B} \mathbf{X}$$

$$I_{\beta\rho_g} = 0; \quad g = 1, 2, \dots, G \quad (\text{A28.a})$$

$$I_{\beta\sigma_{ij}} = 0; \quad i, j = 1, 2, \dots, G$$

$$I_{\rho_g \rho_g} = \text{Tr} \left[\mathbf{B}_g^{-1} \mathbf{W} \mathbf{B}_g^{-1} \mathbf{W}' \right] + \text{tr} \mathbf{B}^{-1} \left[I_T \otimes \left(\mathbf{E}^{gg} \boldsymbol{\Sigma}^{-1} \mathbf{E}^{gg} \right) \otimes \left(\mathbf{W}' \mathbf{W} \right) \right] \mathbf{B}^{-1} \boldsymbol{\Omega}; \quad g = 1, 2, \dots, G$$

$$I_{\rho_g \rho_s} = \text{tr} \mathbf{B}^{-1} \left[I_T \otimes \left(\mathbf{E}^{ss} \boldsymbol{\Sigma}^{-1} \mathbf{E}^{gg} \right) \otimes \left(\mathbf{W}' \mathbf{W} \right) \right] \mathbf{B}^{-1} \boldsymbol{\Omega}; \quad g, s = 1, 2, \dots, G \quad (\text{A28.b})$$

$$I_{\rho_g \sigma_{ij}} = \text{tr} \left[I_T \otimes \left(\mathbf{E}^{gg} \boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \right) \otimes \mathbf{W} \right] \mathbf{B}^{-1}; \quad g; i, j = 1, 2, \dots, G$$

$$I_{\sigma_{ij} \sigma_{sr}} = \frac{\text{TR}}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1} \mathbf{E}^{sr} \right]; \quad i, j, s, r = 1, 2, \dots, G \quad (\text{A28.c})$$

We use the following ordering of the information matrix:

$$I(\theta) = \begin{bmatrix} I_{\beta\beta} & I_{\beta\rho} & I_{\beta\sigma} \\ (k \times k) & (k \times G) & (k \times (G(G+1)/2)) \\ I_{\rho\rho} & I_{\rho\sigma} \\ (G \times G) & (G \times (G(G+1)/2)) \\ I_{\sigma\sigma} \\ ((G(G+1)/2) \\ x(G(G+1)/2)) \end{bmatrix} \quad (\text{A29})$$

The null hypothesis that we want to test is that there is no SEM structure in the error terms of the SUR:

$$\left. \begin{array}{l} H_0: \rho_g = 0; \forall g \\ H_A: \text{No } H_0 \end{array} \right\} \Rightarrow \text{Under } H_0 \Rightarrow \mathbf{B} = I_{\text{TRG}} \Rightarrow \left. \begin{array}{l} y = \mathbf{X}\beta + \varepsilon \\ \varepsilon \sim N(0, \boldsymbol{\Omega}); \boldsymbol{\Omega} = I_T \otimes \boldsymbol{\Sigma} \otimes I_R \end{array} \right\} \quad (\text{A30})$$

The score of (A26) under the null hypothesis of (A30) becomes:

$$\mathbf{g}(\theta)_{|H_0} = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \rho_g} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} \mathbf{g}(\beta)_{|H_0} \\ \mathbf{g}(\rho)_{|H_0} \\ \mathbf{g}(\sigma)_{|H_0} \end{bmatrix} = \begin{bmatrix} 0 \\ \hat{\mathbf{u}}' \left[I_T \otimes \left(\boldsymbol{\Sigma}^{-1} \mathbf{E}^{gg} \right) \otimes \mathbf{W} \right] \hat{\mathbf{u}} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \sum_{t=1}^T \sum_{s=1}^G \sigma^{sg} \hat{\mathbf{u}}_{tg}' \mathbf{W} \hat{\mathbf{u}}_{ts} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \sum_{s=1}^G \sigma^{sg} \text{tr}(\hat{\mathbf{u}}_g' \hat{\mathbf{u}}_{Ls}) \\ 0 \end{bmatrix} \quad (\text{A31})$$

$\hat{\mathbf{u}}$ is the (TRGx1) vector of residuals of the SUR model without spatial effects and $\hat{\mathbf{u}}_L$ its spatial lag. The elements of the information matrix, under the null hypothesis, are:

$$I_{\beta\beta} = \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X}$$

$$I_{\beta\rho_g} = 0; \quad g = 1, 2, \dots, G \quad (\text{A32.a})$$

$$I_{\beta\sigma_{ij}} = 0; \quad i, j = 1, 2, \dots, G$$

$$I_{\rho_g \rho_g} = \text{Ttr}(\mathbf{W}'\mathbf{W})(\sigma^{gg}\sigma_{gg} + 1); \quad g = 1, 2, \dots, G$$

$$I_{\rho_g \rho_s} = T\sigma^{gs}\sigma_{gs}\text{tr}(\mathbf{W}'\mathbf{W}); \quad g, s = 1, 2, \dots, G \quad (\text{A32.b})$$

$$I_{\rho_g \sigma_{ij}} = 0; \quad g; i, j = 1, 2, \dots, G$$

$$I_{\sigma_{ij}\sigma_{sr}} = \frac{\text{TR}}{2} \text{tr}[\boldsymbol{\Sigma}^{-1}\mathbf{E}^{ij}\boldsymbol{\Sigma}^{-1}\mathbf{E}^{sr}]; \quad i, j, s, r = 1, 2, \dots, G \quad (\text{A32.c})$$

The structure of that matrix is, once again, block-diagonal:

$$I(\theta)|_{H_0} = \begin{bmatrix} I_{\beta\beta} & 0 & 0 \\ (k \times k) & (k \times G) & (k \times G) \\ I_{\rho\rho} & 0 & 0 \\ (G \times G) & (G \times (G(G-1)/2)) & (G \times (G(G-1)/2)) \\ I_{\sigma\sigma} & & \\ ((G(G-1)/2) & & \\ \times (G(G-1)/2)) & & \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \left\{ \begin{array}{l} M_{11} = \begin{bmatrix} I_{\beta\beta} & 0 \\ (k \times k) & \\ I_{\sigma\sigma} & \\ ((G(G-1)/2) & \\ \times (G(G-1)/2)) & \end{bmatrix} \\ M_{22} = \begin{bmatrix} I_{\rho\rho} \\ (G \times G) \end{bmatrix} \\ M_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ M_{21} = \begin{bmatrix} 0 & 0 \end{bmatrix} \end{array} \right. \quad (\text{A33})$$

This result facilitates the obtaining of the corresponding Multiplier:

$$\begin{aligned} \mathbf{LM}_{\text{SEM}}^{\text{SUR}} &= \left[\mathbf{g}(\theta)|_{H_0} \right]' \left[I(\theta)|_{H_0} \right]^{-1} \left[\mathbf{g}(\theta)|_{H_0} \right]_{\text{as}} \sim \chi^2(G) \\ \Rightarrow \mathbf{LM}_{\text{SEM}}^{\text{SUR}} &= \mathbf{g}'_{(\rho)|_{H_0}} \left[I_{\rho\rho} \right]^{-1} \mathbf{g}_{(\rho)|_{H_0}} = \\ &= \left[\sum_{t=1}^T \sum_{s=1}^G \sum_{g=1, \dots, G} \sigma^{sg} \hat{u}'_{tg} \mathbf{W} \hat{u}_{ts} \right]' \left[I_{\rho\rho} \right]^{-1} \left[\sum_{t=1}^T \sum_{s=1}^G \sum_{g=1, \dots, G} \sigma^{sg} \hat{u}'_{tg} \mathbf{W} \hat{u}_{ts} \right]_{\text{as}} \sim \chi^2(G) \end{aligned} \quad (\text{A34})$$

Section A.IV. Testing for the constancy of the spatial dependence coefficients.

The log-likelihood function of the SARAR model specified in expression (35) is similar to that of expression (2). The first corresponds to the restricted version and the second to the unrestricted version of the model. The log-likelihood function of the restricted model is:

$$l(y; \theta) = -\frac{\text{RTG}}{2} \ln(2\pi) - \frac{\text{TR}}{2} \ln|\boldsymbol{\Sigma}| + \text{TG} \ln|\hat{\mathbf{B}}| + \text{TG} \ln|\hat{\mathbf{A}}| - \frac{(\mathbf{A}y - \mathbf{X}\beta)' \left[I_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}' \hat{\mathbf{B}}) \right] (\mathbf{A}y - \mathbf{X}\beta)}{2} \quad (\text{A35})$$

The score vector is more compact in the restricted model:

$$\mathbf{g}(\theta) = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \lambda} \\ \frac{\partial l}{\partial \rho} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}' \hat{\mathbf{B}}) \right] (\mathbf{A}y - \mathbf{X}\beta) \\ -\text{TGtr} \left[\hat{\mathbf{A}}^{-1} \mathbf{W} \right] + (\mathbf{A}y - \mathbf{X}\beta)' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}' \hat{\mathbf{B}}) \right] (\mathbf{I}_{\text{TG}} \otimes \mathbf{W}) y \\ -\text{TGtr} \left[\hat{\mathbf{B}}^{-1} \mathbf{W} \right] + (\mathbf{A}y - \mathbf{X}\beta)' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}' \mathbf{W}) \right] (\mathbf{A}y - \mathbf{X}\beta) \\ -\frac{\text{TR}}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \right] + \frac{(\mathbf{A}y - \mathbf{X}\beta)' \left[\mathbf{I}_T \otimes (\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1}) \otimes (\hat{\mathbf{B}}' \hat{\mathbf{B}}) \right] (\mathbf{A}y - \mathbf{X}\beta)}{2} \end{bmatrix} \quad (\text{A36})$$

It is easy to test for the absence of spatial effects in the specification of (35). Now the null hypothesis only affects to two parameters:

$$\left. \begin{array}{l} H_0 : \lambda = \rho = 0 \\ H_A : \text{No } H_0 \end{array} \right\} \Rightarrow \text{Under } H_0 \Rightarrow \mathbf{A} = \mathbf{B} = \mathbf{I}_{\text{TRG}} \quad (\text{A37})$$

The associated Multiplier is:

$$\Rightarrow \mathbf{LM}_{\text{SARAR}}^{\text{SUR(cons)}} = \begin{bmatrix} \mathbf{g}'(\lambda)_{|H_0} & \mathbf{g}'(\rho)_{|H_0} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{\lambda\lambda} - \mathbf{I}_{\lambda\beta} \mathbf{I}_{\beta\beta}^{-1} \mathbf{I}_{\beta\lambda} & \mathbf{I}_{\lambda\rho} \\ \mathbf{I}_{\rho\lambda} & \mathbf{I}_{\rho\rho} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}(\lambda)_{|H_0} \\ \mathbf{g}(\rho)_{|H_0} \end{bmatrix} \underset{\text{as}}{\sim} \chi^2(2) \quad (\text{A38})$$

with:

$$\begin{aligned} \mathbf{I}_{\beta\beta} &= \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R \right] \mathbf{X} \\ \mathbf{I}_{\beta\lambda} &= \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes \mathbf{W} \right] \mathbf{X}\beta \\ \mathbf{I}_{\lambda\lambda} &= \text{TG} \left[\text{tr}(\mathbf{W}' \mathbf{W}) + \text{tr}(\mathbf{W}\mathbf{W}) \right] + \beta' \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\mathbf{W}' \mathbf{W}) \right] \mathbf{X}\beta \\ \mathbf{I}_{\rho\rho} &= \text{TG} \left[\text{tr}(\mathbf{W}' \mathbf{W}) + \text{tr}(\mathbf{W}\mathbf{W}) \right] \\ \mathbf{I}_{\rho\lambda} &= \text{TG} \left[\text{tr}(\mathbf{W}' \mathbf{W}) + \text{tr}(\mathbf{W}\mathbf{W}) \right] \end{aligned} \quad \begin{aligned} \mathbf{g}(\lambda)_{|H_0} &= \sum_{t=1}^T \hat{u}_t' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R) \mathbf{W} y_t \\ \mathbf{g}(\rho)_{|H_0} &= \sum_{t=1}^T \hat{u}_t' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R) \mathbf{W} \hat{u}_t \end{aligned} \quad (\text{A39})$$

The discussion for the Spatial Lag Model of (36) is very similar. Now, the log-likelihood function and the score vector are as follows:

$$l(y; \theta) = -\frac{\text{RTG}}{2} \ln(2\pi) - \frac{\text{TR}}{2} \ln |\boldsymbol{\Sigma}| + \text{TG} \ln |\hat{\mathbf{A}}| - \frac{(\mathbf{A}y - \mathbf{X}\beta)' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R \right] (\mathbf{A}y - \mathbf{X}\beta)}{2} \quad (\text{A40})$$

$$\mathbf{g}(\theta) = \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \lambda} \\ \frac{\partial l}{\partial \rho} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R \right] (\mathbf{A}y - \mathbf{X}\beta) \\ -\text{TGtr} \left[\hat{\mathbf{A}}^{-1} \mathbf{W} \right] + (\mathbf{A}y - \mathbf{X}\beta)' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R \right] (\mathbf{I}_{\text{TG}} \otimes \mathbf{W}) y \\ -\frac{\text{TR}}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \right] + \frac{(\mathbf{A}y - \mathbf{X}\beta)' \left[\mathbf{I}_T \otimes (\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1}) \otimes \mathbf{I}_R \right] (\mathbf{A}y - \mathbf{X}\beta)}{2} \end{bmatrix} \quad (\text{A41})$$

The hypothesis that parameter λ is zero leads us back to the SUR without spatial effects:

$$\left. \begin{array}{l} H_0 : \lambda = 0 \\ H_A : \text{No } H_0 \end{array} \right\} \Rightarrow \text{Under } H_0 \Rightarrow \mathbf{A} = \mathbf{I}_{\text{TRG}} \quad (\text{A42})$$

and the Multiplier results to be:

$$\Rightarrow \mathbf{LM}_{\text{SLM}}^{\text{SUR(cons)}} = \mathbf{g}'(\lambda)_{|H_0} \begin{bmatrix} \mathbf{I}_{\lambda\lambda} - \mathbf{I}_{\lambda\beta} \mathbf{I}_{\beta\beta}^{-1} \mathbf{I}_{\beta\lambda} \end{bmatrix}^{-1} \mathbf{g}(\lambda)_{|H_0} \underset{\text{as}}{\sim} \chi^2(1) \quad (\text{A43})$$

with:

$$\begin{aligned} \mathbf{g}_{(\lambda)_{H_0}} &= \sum_{t=1}^T \hat{\mathbf{u}}_t' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R) \mathbf{W} \mathbf{y}_t \\ \begin{cases} \mathbf{I}_{\beta\beta} = \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R \right] \mathbf{X} \\ \mathbf{I}_{\beta\lambda} = \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R \right] \mathbf{X} \boldsymbol{\beta} \\ \mathbf{I}_{\lambda\lambda} = \text{TG} [\text{tr}(\mathbf{W}' \mathbf{W}) + \text{tr}(\mathbf{W} \mathbf{W})] + \boldsymbol{\beta}' \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\mathbf{W}' \mathbf{W}) \right] \mathbf{X} \boldsymbol{\beta} \end{cases} \end{aligned} \quad (\text{A44})$$

Finally, the log-likelihood function and the score vector for the SEM of (37) are:

$$l(y; \theta) = -\frac{\text{RTG}}{2} \ln(2\pi) - \frac{\text{TR}}{2} \ln |\boldsymbol{\Sigma}| + \text{TG} \ln |\hat{\mathbf{B}}| - \frac{(y - \mathbf{X}\boldsymbol{\beta})' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}' \hat{\mathbf{B}}) \right] (y - \mathbf{X}\boldsymbol{\beta})}{2} \quad (\text{A45})$$

$$\mathbf{g}(\theta) = \begin{bmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}} \\ \frac{\partial l}{\partial \rho_t} \\ \frac{\partial l}{\partial \sigma_{ij}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}' \hat{\mathbf{B}}) \right] (y - \mathbf{X}\boldsymbol{\beta}) \\ -\text{TG} \text{tr} \left[\hat{\mathbf{B}}^{-1} \mathbf{W} \right] + (y - \mathbf{X}\boldsymbol{\beta})' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}' \mathbf{W}) \right] (y - \mathbf{X}\boldsymbol{\beta}) \\ -\frac{\text{TR}}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \right] + \frac{(y - \mathbf{X}\boldsymbol{\beta})' \left[\mathbf{I}_T \otimes (\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1}) \otimes (\hat{\mathbf{B}}' \hat{\mathbf{B}}) \right] (y - \mathbf{X}\boldsymbol{\beta})}{2} \end{bmatrix} \quad (\text{A46})$$

The null hypothesis of absence of spatial effects implies that parameter ρ is zero:

$$\left. \begin{array}{l} H_0: \rho = 0 \\ H_A: \text{No } H_0 \end{array} \right\} \Rightarrow \text{Under } H_0 \Rightarrow \mathbf{B} = \mathbf{I}_{\text{TRG}} \quad (\text{A47})$$

The expression of the Multiplier is:

$$\mathbf{LM}_{\text{SEM}}^{\text{SUR}(\text{cons})} = \mathbf{g}'_{(\rho)_{H_0}} \mathbf{I}_{\rho\rho}^{-1} \mathbf{g}_{(\rho)_{H_0}} \sim \chi^2(1) \quad (\text{A48})$$

$$\begin{aligned} \mathbf{g}_{(\lambda)_{H_0}} &= \sum_{t=1}^T \hat{\mathbf{u}}_t' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_R) \mathbf{W} \hat{\mathbf{u}}_t \\ \mathbf{I}_{\rho\rho} &= \text{TG} [\text{tr}(\mathbf{W}' \mathbf{W}) + \text{tr}(\mathbf{W} \mathbf{W})] \end{aligned}$$

The problem addressed in Section 5.1 refers to the assumption of constancy of the parameters of spatial dependence introduced in the SUR specifications. In the case of the SARAR model associated to the hypothesis of (38), the Lagrange Multiplier appears in expression (39) and the information matrix that should be introduced in this expression is:

$$\mathbf{I}(\theta) = \begin{bmatrix} \mathbf{I}_{\beta\beta} & \mathbf{I}_{\beta\lambda} & \mathbf{I}_{\beta\rho} & \mathbf{I}_{\beta\sigma} \\ & \mathbf{I}_{\lambda\lambda} & \mathbf{I}_{\lambda\rho} & \mathbf{I}_{\lambda\sigma} \\ & & \mathbf{I}_{\rho\rho} & \mathbf{I}_{\rho\sigma} \\ & & & \mathbf{I}_{\sigma\sigma} \end{bmatrix} \quad (\text{A49})$$

where $\hat{u} = [I_T \otimes I_G \otimes (I_R - \hat{\lambda}W)]y - X\hat{\beta}$ and $\hat{\varepsilon} = [I_T \otimes I_G \otimes (I_R - \hat{\rho}W)]\hat{u}$. Furthermore:

$$\begin{cases}
 I_{\beta\beta} = X' [I_T \otimes \Sigma^{-1} \otimes (\hat{B}'\hat{B})] X \\
 I_{\beta\lambda_g} = X' [I_T \otimes (\Sigma^{-1}E^{gg}) \otimes (\hat{B}'W\hat{B}\hat{A}^{-1})] X\beta; \quad g=1,2,\dots,G \\
 I_{\beta\rho_g} = 0; \quad g=1,2,\dots,G \\
 I_{\beta\sigma_{ij}} = 0; \quad i,j=1,2,\dots,G \\
 I_{\lambda_g\lambda_g} = T\text{Gtr}(\hat{A}'^{-1}W'\hat{A}^{-1}W) + \sigma^{gg}\beta'X' [I_T \otimes E^{gg} \otimes \hat{A}'^{-1}W'\hat{B}'\hat{B}W\hat{A}^{-1}] X\beta \\
 \quad + T\sigma^{gg}\sigma_{gg}\text{tr} \left[\hat{A}'^{-1}W'\hat{B}'\hat{B}\hat{A}^{-1}W(\hat{B}'\hat{B})^{-1} \right] \\
 I_{\lambda_g\lambda_s} = \sigma^{gs}\beta'X' [I_T \otimes E^{gs} \otimes (\hat{A}'^{-1}W'\hat{B}'\hat{B}W\hat{A}^{-1})] X\beta \\
 \quad + T\sigma^{gs}\sigma_{gs}\text{tr} \left[\hat{A}'^{-1}W'\hat{B}'\hat{B}\hat{A}^{-1}W(\hat{B}'\hat{B})^{-1} \right]; \quad g,s=1,2,\dots,G \\
 I_{\lambda_g\rho_g} = T\sigma^{gg}\sigma_{gg}\text{tr} \left[\hat{B}'^{-1}W'\hat{B}W\hat{A}^{-1}\hat{B}^{-1} \right] + T\text{tr}(WW\hat{A}^{-1}\hat{B}^{-1}); \quad g=1,2,\dots,G \\
 I_{\lambda_g\rho_s} = T\sigma^{gs}\sigma_{gs}\text{tr} \left[\hat{B}'^{-1}W'\hat{B}W\hat{A}^{-1}\hat{B}^{-1} \right]; \quad g,s=1,2,\dots,G \\
 I_{\lambda_g\sigma_{ij}} = \begin{cases} 0 & \text{if } t \neq i \text{ or } g \neq j \\ T\sigma^{gg}\text{tr}(\hat{A}^{-1}W) & \text{if } g=i=j \end{cases} \\
 I_{\rho_g\rho_g} = T\text{Gtr}[\hat{B}'^{-1}W\hat{B}^{-1}W] + T\sigma^{gg}\sigma_{gg}\text{tr}[\hat{B}'^{-1}WW'\hat{B}^{-1}]; \quad g=1,2,\dots,G \\
 I_{\rho_g\rho_s} = T\sigma^{gs}\sigma_{gs}\text{tr}[\hat{B}'^{-1}WW'\hat{B}^{-1}]; \quad g,s=1,2,\dots,G \\
 I_{\rho_g\sigma_{ij}} = \begin{cases} 0 & \text{if } g \neq i \text{ and } g \neq j \\ T\sigma^{gi}\text{tr}[\hat{B}'^{-1}W] & \text{if } g=i \text{ or } t=j \end{cases} \\
 I_{\sigma_{ij}\sigma_{sr}} = \frac{\text{TR}}{2} \text{tr}[\Sigma^{-1}E^{ij}\Sigma^{-1}E^{sr}]; \quad i,j,s,r=1,2,\dots,G
 \end{cases} \tag{A50}$$

In the case of the SLM model to which refers the hypothesis of (41), we will need the following information matrix in order to solve for the expression of the Lagrange Multiplier that appears in (42):

$$\begin{cases}
 I^{T,R}(\theta) = \begin{bmatrix} I_{\beta\beta} & I_{\beta\lambda} & I_{\beta\sigma} \\ & I_{\lambda\lambda} & I_{\lambda\sigma} \\ & & I_{\sigma\sigma} \end{bmatrix} \\
 I_{\beta\beta} = X' [I_T \otimes \Sigma^{-1} \otimes I_R] X \\
 I_{\beta\lambda_g} = X' [I_T \otimes (\Sigma^{-1}E^{gg}) \otimes (W\hat{A}^{-1})] X\beta; \quad g=1,2,\dots,G \\
 I_{\beta\sigma_{ij}} = 0; \quad i,j=1,2,\dots,T \\
 I_{\lambda_g\lambda_g} = T\text{Gtr}(\hat{A}'^{-1}W'\hat{A}^{-1}W) + \sigma^{gg}\beta'X' [I_T \otimes E^{gg} \otimes \hat{A}'^{-1}W'W\hat{A}^{-1}] X\beta \\
 \quad + T\sigma^{gg}\sigma_{gg}\text{tr}[\hat{A}'^{-1}W'W\hat{A}^{-1}]; \quad g=1,2,\dots,G \\
 I_{\lambda_g\lambda_s} = \sigma^{gs}\beta'X' [I_T \otimes E^{gs} \otimes (\hat{A}'^{-1}W'W\hat{A}^{-1})] X\beta + T\sigma^{gs}\sigma_{gs}\text{tr}[\hat{A}'^{-1}W'W\hat{A}^{-1}] \\
 \quad ; t,s=1,2,\dots,G \\
 I_{\lambda_g\sigma_{ij}} = \begin{cases} 0 & \text{if } g \neq i \text{ or } g \neq j \\ T\sigma^{gg}\text{tr}(\hat{A}^{-1}W) & \text{if } t=i=j \end{cases} \\
 I_{\sigma_{ij}\sigma_{sr}} = \frac{\text{TR}}{2} \text{tr}[\Sigma^{-1}E^{ij}\Sigma^{-1}E^{sr}]; \quad i,j,s,r=1,2,\dots,G
 \end{cases} \tag{A51}$$

Finally, the same can be said with respect to the SEM model of (10). The null hypothesis of time-constancy appears in (24), the Lagrange Multiplier is that of (45) in which intervenes the following information matrix:

$$\begin{aligned}
 \mathbf{I}(\theta) &= \begin{bmatrix} I_{\beta\beta} & I_{\beta\rho} & I_{\beta\sigma} \\ & I_{\rho\rho} & I_{\rho\sigma} \\ & & I_{\sigma\sigma} \end{bmatrix} \\
 \left\{ \begin{aligned} I_{\beta\beta} &= \mathbf{X}' \left[\mathbf{I}_T \otimes \boldsymbol{\Sigma}^{-1} \otimes (\hat{\mathbf{B}}' \hat{\mathbf{B}}) \right] \mathbf{X} \\ I_{\beta\rho_t} &= 0; \quad t = 1, 2, \dots, G \\ I_{\beta\sigma_{ij}} &= 0; \quad i, j = 1, 2, \dots, G \\ I_{\rho_g \rho_g} &= T \text{Gtr} \left[\hat{\mathbf{B}}'^{-1} \mathbf{W} \hat{\mathbf{B}}^{-1} \mathbf{W} \right] + T \sigma^{gg} \sigma_{gg} \text{tr} \left[\hat{\mathbf{B}}'^{-1} \mathbf{W}' \mathbf{W} \hat{\mathbf{B}}^{-1} \right]; \quad g = 1, 2, \dots, G \\ I_{\rho_g \rho_s} &= T \sigma^{gs} \sigma_{gs} \text{tr} \left[\hat{\mathbf{B}}'^{-1} \mathbf{W} \mathbf{W}' \hat{\mathbf{B}}^{-1} \right]; \quad g, s = 1, 2, \dots, G \\ I_{\rho_g \sigma_{ij}} &= \begin{cases} 0 & \text{if } g \neq i, g \neq j \text{ and } i \neq j \\ T \sigma^{gi} \text{tr} \left[\hat{\mathbf{B}}'^{-1} \mathbf{W} \right] & \text{if } i \neq j \text{ and } g=i \text{ or } t=j \\ T \sigma^{gg} \text{tr} \left[\hat{\mathbf{B}}'^{-1} \mathbf{W} \right] & \text{if } g=i=j \end{cases} \\ I_{\sigma_{ij} \sigma_{sr}} &= \frac{TR}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \mathbf{E}^{ij} \boldsymbol{\Sigma}^{-1} \mathbf{E}^{sr} \right]; \quad i, j, s, r = 1, 2, \dots, G \end{aligned} \right. \quad (\text{A52})
 \end{aligned}$$