

**WORKING PAPERS**

N° 1192

February 2021

# “Revisiting Estimation Methods for Spatial Econometric Interaction Models”

Lukas Dargel

# Revisiting Estimation Methods for Spatial Econometric Interaction Models

Lukas Dargel 

Toulouse School of Economics, University of Toulouse Capitole, France

February 24, 2021

## Abstract

Taking advantage of a generalization of the matrix formulation introduced by LeSage and Pace (2008), this article presents improvements in the computational performance and flexibility of three estimators of spatial econometric interaction models. By generalizing computational techniques for the evaluation of the likelihood function and also for the Hessian matrix the maximum likelihood estimator (MLE) achieves computation times that are not much longer than those of an ordinary least-squares (OLS) regression. The restructured likelihood also improves the performance of the Bayesian Markov chain Monte Carlo (MCMC) estimator considerably. Finally, the spatial two-stage least-squares (S2SLS) estimator presented in this article is the first one that exploits the efficiency gains of the matrix formulation. In addition to the computational improvements of the three estimation methods this article presents a new solution to the issue of defining the feasible parameter space that allows to verify the consistency of the spatial econometric interaction model with a minimal computational burden. All of these developments indicate that the spatial econometric alternative to the traditional gravity model has become an increasingly mature option and should eventually be considered a standard modeling approach for origin-destination flow problems.

**Keywords:** Origin-destination flows, Cross-sectional dependence, Maximum likelihood, Two-stage least-squares, Bayesian Markov chain Monte Carlo

**JEL Codes:** C01, C21, C63

# 1 Introduction

Spatial interaction models are a mathematical representation of interaction behavior between entities located at an origin and a destination. They are widely applied to explain and predict, for example, passenger-flows between airports, trade or migration flows between countries, and flows of customers from residential areas to stores. The most famous spatial interaction model is the gravity model, which is originally inspired by Newton’s law of universal gravitation. The gravity model states that the attraction between an origin and a destination increases as their masses increase and the distance between them decreases. Young (1924) was among the first to use a simple gravity model to explain migration flows of the agricultural population between farms. Since then, the gravity model has been continuously refined and has become one of the most popular econometric models used across industries and scientific disciplines. A recent overview of the topic is given by Oshan (2020).

Initially, the success of the gravity model has been solely attributed to its excellent agreement with the data in empirical applications. However, by taking a macro perspective of interactions between cities or countries, it lacked a solid theoretical underpinning. Wilson (1967) addressed this problem by showing that the gravity equation can be derived from the behavior of individual actors. This theoretical underpinning increased the recognition of the model in academic circles, but Curry (1972) raised another type of criticism when she showed that the distance parameter is not estimated consistently by usual OLS estimation. This would only be possible if the flows were independent, an assumption that is rather unrealistic as many flows start from the same origin or go to the same destination. Additionally, an increasing number of studies report significant spatial dependence in origin-destination flows (see for example Lee and Pace 2005; Porojan 2001), which compromises not only the computation of the distance effect but the estimation of all model parameters.

LeSage and Pace (2008) account for spatial dependence in gravity type models by extending the class of spatial econometric models to origin-destination flows. In general, spatial econometric models avoid the independence assumption by allowing observations to be influenced by their geographical neighbors. The neighborhood structure is reflected by spatial weight matrices which allow to compute spatial lags of the original variables. These spatial lags contain information on the average value of a variable in the neighborhood around a location and a variety of spatial econometric models use them to account for different types of spatial dependence. A great overview on different types of spatial econometric models is provided by Elhorst (2010). The spatial lag (LAG) model includes a lag of the dependent variable and using additional lags of the explanatory variables leads to the so-called Spatial Durbin model (SDM). These models cannot be estimated consistently by OLS and much effort has been devoted to the development of alternative estimation procedures, such as MLE (Ord 1975; Anselin 1988), S2SLS estimation (Kelejian and Prucha 1998), or Bayesian MCMC (LeSage 1997). Naturally, these procedures have much higher computational cost than the OLS method and a large strand of the literature is devoted to improvements that bring down the cost of estimating spatial econometric models. For interaction models the usual computational bottlenecks are even more pronounced because they have a double spatial index. This double index complicates the definition of a geographical neighborhood because a flow might be influenced by the neighbors of the origin, the destination, or both. Additionally, the number of potential interactions  $N = n^2$  grows quadratically with the number of origins and destinations, which increases problems related to the dimension of the data. LeSage and Pace (2008) propose a spatial interaction model that represents the spatial dependence structure of the flows with three neighborhood matrices. They also attenuate the problem of high dimensions by developing a matrix formulation of the model, that is much more efficient than the traditional vectorized formulation.

This article further improves some computational aspects related to the estimation of the

spatial econometric interaction model and puts forward a generalization of the matrix formulation that allows to use the exogenous variables much more flexibly. The MLE and the Bayesian MCMC estimator benefit from a simple reformulation of the likelihood function that allows to compute the residual sum of squares (RSS) term from moment matrices with low dimensions. We also extend the mixed numerical analytical Hessian approximation of LeSage and Pace (2009) to higher order spatial models and show how to use this extension to compute the variances of the MLE at a very low cost. To improve the performance of the S2SLS estimator we develop its matrix formulation and also address multicollinearity problems of the spatial instruments. Some of these problems arise as a consequence of the origin-destination structure of the data and others are linked to the SDM specification of the model. Consequently, the presented solutions to the multicollinearity of instruments in the SDM also help when estimating spatial models that do not deal with flows. Concerning the feasible parameter space, we develop a new method that allows to verify the consistency of the model and only marginally increases the computational burden of the estimation. To facilitate the adoption of the model all estimators presented in this article are implemented in the R package `spflow`. These contributions should help to establish spatial econometric interaction models as a standard approach to modeling origin-destination flows. Failing to make this model operational implies that spatial dependence continues to be ignored in many studies which may have severe consequences. Kerkman, Martens, and Meurs (2017), for example, point out that omitted spatial autocorrelation might be a reason for the systematic overestimation of the effects of policy interventions in the public transport sector.

Section 2 introduces the spatial model and the matrix formulation. Section 3 discusses the issue of defining the feasible parameter space. Section 4 details the improved calculation procedures for the three estimators. In Section 5 all estimators are illustrated based on the example of home-to-work commuting flows between the municipalities of Paris. The final section concludes.

## 2 A spatial econometric model for origin-destination flows

LeSage and Pace (2008) develop a model for origin-destination flows that accounts for three types of spatial autocorrelation. Their matrix formulation leads to significant efficiency gains as it allows to perform most calculations linked to the parameter estimation with objects of dimension  $n$  (the number of spatial sites) instead of  $N = n^2$  (the number of origin-destination pairs). The following paragraphs introduce this model and present a generalized version of the matrix formulation that is more flexible in handling the explanatory variables.

### Representation of the flows

To construct the flow matrix  $Y(n \times n)$  we have to define an ordering of all the spatial sites at the origins and destinations of the flows. Based on this ordering, all flows are arranged into a matrix, whose columns correspond to the origins and whose rows correspond to the destinations, as is shown in (1) below.

$$\begin{pmatrix} o_1 \rightarrow d_1 & o_2 \rightarrow d_1 & \cdots & o_n \rightarrow d_1 \\ o_1 \rightarrow d_2 & o_2 \rightarrow d_2 & \cdots & o_n \rightarrow d_2 \\ \vdots & \vdots & \ddots & \vdots \\ o_1 \rightarrow d_n & o_2 \rightarrow d_n & \cdots & o_n \rightarrow d_n \end{pmatrix} \Rightarrow \begin{pmatrix} y_1 & y_{n+1} & \cdots & y_{N-n+1} \\ y_2 & y_{n+2} & \cdots & y_{N-n+2} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & y_{2n} & \cdots & y_N \end{pmatrix} = Y \quad (1)$$

Using the VEC-operator that stacks the columns of a matrix (from left to right), it is then possible to derive the flow vector as  $y = \text{VEC}(Y) = (y_1, y_2, \dots, y_N)'$ . LeSage and Pace (2009)

describe this representation of  $y$  as an origin-centric flow arrangement. Applying the same procedure to the transpose  $Y'$  of the flow matrix would lead to a destination centric flow arrangement.

## Representation of the spatial neighborhood

To illustrate how LeSage and Pace (2008) derive the three neighborhood matrices of the origin-destination pairs we consider a toy example of three sites ( $s_1, s_2, s_3$ ) that are located on a line. In this example the neighborhood matrix  $W$  of the sites is defined by contiguity.

$$\begin{array}{|c|} \hline s_1 \\ \hline s_2 \\ \hline s_3 \\ \hline \end{array} \implies W = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} \quad (2)$$

By convention the diagonal elements of  $W$  are zero which means that no site is its own neighbor. We also make the assumption that the rows of  $W$  sum to one, which is common in the spatial econometric literature. This row-stochastic property yields some computational advantages and allows to interpret the spatial lag of a variable as a local weighted average, where the weights correspond to the strength of the neighborhood relation. Based on these assumptions, we obtain the site neighborhood matrix shown in (2). The first and last sites ( $s_1$  and  $s_3$ ) only have the middle site ( $s_2$ ) as their neighbor, which in turn has two neighbors. From the site neighborhood matrix  $W$  we can then derive three origin-destination pair neighborhood matrices

$$W_d = \mathbf{I}_n \otimes W \quad W_o = W \otimes \mathbf{I}_n \quad W_w = W \otimes W, \quad (3)$$

where  $\otimes$  denotes the Kronecker product. The destination-neighborhood matrix  $W_d$  allows to capture the mutual influence of flows that start at the same origin and end at a neighbor of the destination. Likewise, the origin-neighborhood matrix  $W_o$  represents the dependence between flows that have the same destination and start from neighboring origins. Finally, the origin-to-destination neighborhood matrix  $W_w$  represents links between origin-destination pairs whose origins and destinations are neighbors. All three matrices are of dimensions ( $N \times N$ ) and inherit the row-stochastic property of  $W(n \times n)$ .

Below we illustrate how the  $W_d$  matrix is constructed for the above example of  $n = 3$  sites and  $N = n^2 = 9$  origin-destination pairs. The first row of  $W_d$  reveals that the only destination neighbor of the first pair ( $s_1 \rightarrow s_1$ ) is the second pair ( $s_1 \rightarrow s_2$ ).

$$\begin{array}{|c|} \hline s_1 \rightarrow s_1 \\ \hline s_1 \rightarrow s_2 \\ \hline s_1 \rightarrow s_3 \\ \hline s_2 \rightarrow s_1 \\ \hline s_2 \rightarrow s_2 \\ \hline s_2 \rightarrow s_3 \\ \hline s_3 \rightarrow s_1 \\ \hline s_3 \rightarrow s_2 \\ \hline s_3 \rightarrow s_3 \\ \hline \end{array} \implies W_d = \begin{pmatrix} W & 0 & 0 \\ 0 & W & 0 \\ 0 & 0 & W \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

## The explanatory variables

Interaction models use two types of exogenous variables which have different dimensions. While attributes of the sites have  $n$  entries, those of the origin-destination pairs have  $N = n^2$  entries.

## Origin-destination pair attributes

In the classical gravity model the origin-destination pair attributes correspond to the distance vector  $g$  ( $N \times 1$ ). LeSage and Pace (2008) arrange this vector into the matrix  $G$  ( $n \times n$ ) using the same ordering as the one for the flow matrix in (1). Modern interaction models often employ much broader measures for separation than the geographical distance and it is also possible to use more than one separation measure. An extension to account for multiple pair attributes is not mathematically involving but leads to more complicated moment expressions. For this reason the article works with a single distance vector while the accompanying R package <sup>1</sup> allows multiple origin-destination pair attributes.

## Site attributes

The attributes of the spatial sites reflect the origin's capacity to emit and the destination's capacity to attract a flow. While the original framework of LeSage and Pace (2008) assumed that emission and attraction arise from identical attributes, we will add the option to use distinct sets of variables for these two effects. We introduce this flexibility by using the matrix  $OX$  ( $k_o \times n$ ) to describe the emission capacity of the origins, and the matrix  $DX$  ( $k_d \times n$ ) to describe the attraction capacity of the destinations.

Another frequent problem during the estimation of spatial interaction models is related to flows that start and end at the same geographical site. Such intra-regional flows are often orders of magnitudes larger than inter-regional ones and may dominate the parameter estimation (see for example Tiefelsdorf 2003). Setting the diagonal elements of the flow matrix to zero or removing them from the data appears as an easy solution but it also impacts all inter-regional flows that are neighbors to intra-regional ones. This spurious side effect occurs as soon as the model includes spatial lags of the flow vector and leads to distorted parameter estimates. LeSage and Pace (2009) present a more elegant solution that isolates the dominating data points by adding an additional set of parameters to intra-regional flows. The variables used for the intra-regional model are represented in a third matrix, denoted by  $IX$  ( $k_I \times n$ ). Following a suggestion of LeSage and Pace (2008), this article uses the notation in (4) below to treat the SDM and the LAG specification in the same framework. The same notation is subsequently applied to all three matrices of site attributes ( $DX$ ,  $OX$  and  $IX$ ).

$$\tilde{O}X = \begin{cases} OX & \text{if spatial structure = LAG} \\ (OX \ WOX) & \text{if spatial structure = SDM} \end{cases} \quad (4)$$

## The matrix of explanatory variables

The matrix  $Z$  gathers all explanatory variables in their vectorized formulation. The first two columns of the  $Z$  matrix represent the global constant  $\iota_N$  and an intra-regional constant  $\iota_I = \text{VEC}(I_n)$ , which is equal to one if the origin and destination coincide and zero otherwise. The next four sets of variables represent the attributes of the origins ( $\tilde{X}_o$ ), the destinations ( $\tilde{X}_d$ ), the intra-regional observations ( $\tilde{X}_I$ ) and the distance vector. To construct the matrices  $\tilde{X}_d$ ,  $\tilde{X}_o$  and  $\tilde{X}_I$  we have to duplicate the information in  $DX$ ,  $\tilde{O}X$  and  $IX$ . The ordering introduced in (1) allows to derive all six components of the matrix  $Z = (\iota_N \ \iota_I \ \tilde{X}_d \ \tilde{X}_o \ \tilde{X}_I \ g)$  as

$$\begin{aligned} \iota_N &= \iota_n \otimes \iota_n & \iota_I &= \text{VEC}(I_n) & g &= \text{VEC}(G) \\ \tilde{X}_d &= \iota_n \otimes DX & \tilde{X}_o &= \tilde{O}X \otimes \iota_n & \tilde{X}_I &= \text{Diag}(\iota_I)(IX \otimes \iota_n), \end{aligned} \quad (5)$$

<sup>1</sup>The package will be made available on the author's GitHub page and a release to the Comprehensive R Archive Network (CRAN) is planned in the second quarter of 2021.

where  $I_n$  is the  $(n \times n)$  identity matrix,  $\iota_n$  is a  $(n \times 1)$  vector of ones,  $\otimes$  denotes the Kronecker product and the operation  $\text{Diag}(x)$  creates a matrix whose diagonal contains the elements of the vector  $x$ , while all other elements are zero.

This definition of the explanatory variables implies that the parameters of intra-regional observations only capture the difference to the corresponding inter-regional ones. In other words, the direct effect of a site attribute on an intra-regional flow corresponds to the sum three coefficients if the attribute is present in all three matrices  $\tilde{D}\tilde{X}$ ,  $\tilde{O}\tilde{X}$  and  $\tilde{I}\tilde{X}$ . An alternative parametrization requires to set the rows of  $\iota_N$ ,  $\tilde{X}_d$  and  $\tilde{X}_o$  for intra-regional observations to zero, which leads to intra-regional parameters that capture the full effect of the intra-regional attributes. However, the parametrization implied by equation (5) simplifies the calculation and reporting of significance tests. Using (5) we can, for example, test the necessity to include an intra-regional effect based on a null-hypothesis of the form  $H_0 : \alpha_I = 0$  (where  $\alpha_I$  could be the coefficient of the intra-regional constant). The corresponding t-test is easy to compute and can be reported for all parameters in tabular form. With the alternative parametrization we would have to use the hypothesis of equality between the intra- and inter-regional parameters (for example  $H_0 : \alpha_I = \alpha$ ) to make the same test, which is not as easy to calculate and to report.

## The model

We consider the following spatial econometric interaction model

$$y = \rho_d W_d y + \rho_o W_o y + \rho_w W_w y + Z\delta + \varepsilon \iff Ay = Z\delta + \varepsilon, \quad (6)$$

where the parameters  $\delta$  capture the influence of the explanatory variables,  $\varepsilon$  represents a vector of random errors, and  $\rho = (\rho_d \quad \rho_o \quad \rho_w)'$  is the vector of autocorrelation parameters associated to the three possible forms of spatial dependence. The spatial filter matrix  $A = A(\rho) = (\mathbf{I}_n - \rho_d W_d - \rho_o W_o - \rho_w W_w)$  defines the spatial structure of the model. Definition (6) of the model is, up the elements of the  $Z$  matrix, identical to the one of LeSage and Pace (2008), who derive a family of nine sub-models from this general specification.

## Matrix formulation and moment expressions

Model (6) explains the flow vector  $y$  ( $N \times 1$ ) with some explanatory variables and three autoregressive components. From a theoretical point of view, it is no problem to treat this model as a usual spatial econometric model with multiple weight matrices. However, this vectorized formulation is computationally inefficient because it ignores redundancies that arise from the fact that every site is at the same time origin and destination of many flows. For this reason, the spatial neighborhood matrices and the explanatory variables contain mainly duplicated information. The matrix formulation of LeSage and Pace (2008) avoids these inefficiencies by operating on the flows in their matrix representation and by deriving the elements of moments expressions such as  $Z'Z$  and  $Z'y$  directly from the original data. Hence, when using this matrix formulation we do not need to construct the large matrix of covariates  $Z$  ( $N \times K$ ) and the even larger neighborhood matrices of the origin-destination pairs  $W_r$  ( $N \times N$ ) (for  $r = d, o, w$ ). LeSage and Pace (2008) use these ideas to derive a matrix based MLE and also apply them to a Bayesian MCMC estimator (LeSage and Pace 2009). In Section 4 we will show that this matrix formulation can also be used for a S2SLS estimator of the model.

Moment expressions such as  $Z'Z$  and  $Z'y$  have the advantage that their dimensions are independent of the number of observations and a key idea of the improvements presented in Section 4 is to derive estimators that can be computed from such moment expressions. The moments presented in the following have the same structure as those of LeSage and Pace (2009, page 224) but do not impose that  $\tilde{O}\tilde{X} = \tilde{D}\tilde{X} = \tilde{I}\tilde{X} = \tilde{X}$ .

The  $Z'Z$  moment is derived from the exogenous covariates defined in (5). Since it is a symmetric matrix, we use the  $\bullet$  notation to indicate elements that are already given on the upper triangle. The following definition of  $Z'Z$  also uses the trace operator ( $\text{tr}$ ), the elementwise Hadamard product  $\odot$  and the diag-operator to extract the main diagonal of a square matrix.

$$Z'Z = \begin{pmatrix} N & n & n \cdot \iota'_n \tilde{D}\tilde{X} & n \cdot \iota'_n \tilde{O}\tilde{X} & \iota'_n \tilde{I}\tilde{X} & \iota'_n G \iota_n \\ \bullet & n & \iota'_n \tilde{D}\tilde{X} & \iota'_n \tilde{O}\tilde{X} & \iota'_n \tilde{I}\tilde{X} & \text{tr}(G) \\ \bullet & \bullet & n \cdot \tilde{D}\tilde{X}'\tilde{D}\tilde{X} & (\iota_n \tilde{D}\tilde{X})' \iota_n \tilde{O}\tilde{X} & \tilde{D}\tilde{X}'\tilde{I}\tilde{X} & \tilde{D}\tilde{X}'G\iota_n \\ \bullet & \bullet & \bullet & n \cdot \tilde{O}\tilde{X}'\tilde{O}\tilde{X} & \tilde{O}\tilde{X}'\tilde{I}\tilde{X} & \tilde{O}\tilde{X}'G'\iota_n \\ \bullet & \bullet & \bullet & \bullet & \tilde{I}\tilde{X}'\tilde{I}\tilde{X} & \tilde{I}\tilde{X}'\text{diag}(G) \\ \bullet & \bullet & \bullet & \bullet & \bullet & \iota'_n (G \odot G) \iota_n \end{pmatrix} \quad (7)$$

The  $Z'y$  moment corresponds to the empirical covariance between the flows and the exogenous variables. This covariance is also required for the three spatial lags of the flow vector. LeSage and Pace (2008) note the similarity in these four covariance terms and by adopting their notation we can derive them efficiently according to the equations in (8) below, where  $y^{(t)}$  ( $t = 1, 2, 3, 4$ ) represent the original flow vector as well as the lagged versions ( $y^{(1)} = y$ ,  $y^{(2)} = W_d y$ ,  $y^{(3)} = W_o y$ ,  $y^{(4)} = W_w y$ ).

$$y^{(t)} = \text{VEC}(Y^{(t)}) \quad Y^{(t)} = \begin{cases} Y & \text{if } t = 1 \\ WY & \text{if } t = 2 \\ YW' & \text{if } t = 3 \\ WYW' & \text{if } t = 4 \end{cases} \quad Z'y^{(t)} = \begin{pmatrix} \iota'_n Y^{(t)} \iota_n \\ \text{tr}(Y^{(t)}) \\ \tilde{D}\tilde{X}' Y^{(t)} \iota_n \\ \tilde{O}\tilde{X}' Y^{(t)} \iota_n \\ \tilde{I}\tilde{X}' \text{diag}(Y^{(t)}) \\ \iota'_n (G \odot Y^{(t)}) \iota_n \end{pmatrix} \quad (8)$$

### 3 The feasible parameter space

Unfortunately, the spatial interaction model in (6) is not defined for all values of the autoregressive parameters and we have to consider a feasible parameter space that ensures the well-behavedness of the model. The following paragraphs give an overview about the problems we encounter when defining this feasible parameter space and discuss some of the available solutions. We treat this issue first from the general perspective of a higher-order spatial LAG or Durbin model. Afterwards, we develop a new solution to this problem for the special case of the spatial interaction model in (6). In general, a spatial model of order  $l$  is characterized by the fact that it uses  $l$  distinct neighborhood matrices associated to  $l$  autoregressive parameters. The spatial filter matrix of such a model is given by  $A = (I - W_F)$ , where  $W_F = \rho_1 W_1 + \dots + \rho_l W_l$ .

Table 1 illustrates how the most commonly applied constraints on the parameter space for a spatial model of order one have been extended to higher order models. The presented formulas and the discussion in the following paragraph are valid under the condition that the neighborhood matrices have positive, real entries. Stricter assumptions such as symmetry are not necessary if we interpret the absolute value of a complex number as the modulus. The first column of Table 1 defines five constraints on the feasible space for the autoregressive parameters  $\rho = (\rho_1, \dots, \rho_l)$  in a SDM or LAG model of order  $l$ . Columns two and three of Table 1 indicate whether a given constraint is necessary or sufficient to ensure that the model is well defined and the last column shows an expression of the feasible parameter space for the order one model. The first four constraints are ordered in decreasing restrictiveness, while the fifth constraint cannot be compared to the others because it is rather an ad hoc rule than a mathematically justified restriction.

Table 1: The feasible parameter space in spatial autoregressive models

	Constraint <sup>a</sup>	Necessary	Sufficient	Constraint for $(l = 1)$ <sup>b</sup>
(I)	$\rho \in \mathbb{R}^l \setminus \mathbb{A}_1$	✓	✓	$\rho_1 \neq \lambda_i(W_1)^{-1}$ , for $i = 1, \dots, n$
(II)	$\rho \in \mathbb{B}_1$	✗	✓	$\lambda_{min}(W_1)^{-1} < \rho_1 < \lambda_{max}(W_1)^{-1}$
(III)	$\rho \in \mathbb{B}_2$	✗	✓	$-\phi(W_1)^{-1} < \rho_1 < \phi(W_1)^{-1}$
(IV)	$0 \leq  \rho_1  + \dots +  \rho_l  < 1$	✗	✓	$-1 < \rho_1 < 1$
(V)	$-1 < \rho_1, \dots, \rho_l, \sum_i \rho_i < 1$	✗	✗	$-1 < \rho_1 < 1$

<sup>a</sup> The set  $\mathbb{A}_1$  contains all values of  $\rho$  for which at least one eigenvalue of  $W_F$  is equal to one.

<sup>a</sup> The set  $\mathbb{B}_1$  contains all values of  $\rho$  for which all real eigenvalues of  $W_F$  are smaller than one.

<sup>a</sup> The set  $\mathbb{B}_2$  contains all values of  $\rho$  for which all eigenvalues of  $W_F$  are smaller than one in absolute value.

<sup>b</sup> Let  $M$  be a square matrix, then  $\lambda_i(M)$  is its  $i$ th eigenvalue,  $\lambda_{min}(M)$  (respectively  $\lambda_{max}(M)$ ) its smallest (respectively largest) real eigenvalue and  $\phi(M)$  is the spectral radius of  $M$ .

<sup>b</sup> Constraint (IV) and (V) are typically used in conjunction with row-stochastic neighborhood matrices which always have 1 as their largest eigenvalue.

The first three constraints in Table 1 can be expressed in terms of the eigenvalues  $\lambda(W_F)$  of  $W_F$ . For the order one model the eigenvalues of  $W_F$  are linked to those of  $W_1$  by the equation  $\lambda(W_F) = \rho_1 \lambda(W_1)$ . This link allows to obtain  $\lambda(W_F)$  for changing values of the autoregressive parameter without having to compute an eigenvalue decomposition of the  $W_F$  matrix. In contrast, for higher order models the eigenvalues of  $W_F$  must be recomputed for every change in the autoregressive parameters, which makes it very costly to define the parameter spaces implied by the first three constraints. This difficulty explains why many empirical studies resort to the naive extensions described by constraint (IV) and (V).

Constraint (I) ensures that the variance-covariance matrix of the dependent variable is positive definite and is a necessary and sufficient condition for the spatial model to be well behaved. Martin (1992) and Kelejian and Robinson (1995)<sup>2</sup> derive this condition for the order one model but their conclusions also apply to the case of an order  $l$  model, as shown by the following arguments. If we consider the errors to be independent and gaussian, the variance of the dependent variable is given by  $\text{Var}(y) = (A'A)^{-1}$ . This expression is positive definite when the inverse of the spatial filter matrix exists, which corresponds to all eigenvalues of  $W_F$  being different from one. Even for the order one model, constraint (I) is rarely used in practice because it requires knowledge of all eigenvalues of  $W_F$  and implies that the set of feasible values for the autoregressive parameters is not compact (Hepple 1995). For these reasons, most spatial econometric models employ one of the more restrictive alternatives given by the later four constraints, as they are easier to verify and have more convenient properties.

The second constraint requires that all real eigenvalues of  $W_F$  are smaller than one, which is equivalent to  $\lambda_{max}(W_F) < 1$ . This restriction also implies that the spatial filter matrix is positive definite if no complex eigenvalue of  $W_F$  exceeds 1 in magnitude. The parameter space associated to constraint (II) corresponds to the largest compact set of feasible parameter values that allows zero autocorrelation and is therefore intuitively appealing. In the order one model, the largest real eigenvalue of  $W_F$  is given by  $\lambda_{min}(W_1) \cdot \rho_1$  if  $\rho_1 \leq 0$  and by  $\lambda_{max}(W_1) \cdot \rho_1$  if  $0 < \rho_1$ . From this definition of  $\lambda_{max}(W_F)$  we can easily derive the simple interval in column four and row two of Table 1. However, when  $l \geq 2$  we can identify the space of feasible combinations of the autoregressive parameters (for which  $\lambda_{max}(W_F) < 1$ ) only by numerical search over a grid of possible values for  $(\rho_1, \dots, \rho_l)$ . Elhorst, Lacombe, and Piras (2012) develop a method that

<sup>2</sup>In fact Kelejian and Robinson (1995) discuss a model with spatial autocorrelation in the error term, but most of their considerations are also relevant to the LAG and SDM model because they have a very similar expression of the variance-covariance matrix of the dependent variable.

allows to identify this space based on a grid with dimension  $l - 1$  instead of  $l$ , but computing  $\lambda_{max}(W_F)$  for all points in this grid is still very costly when  $l \geq 3$  and often impractical for large data sets.

The third constraint is more restrictive than the second and requires all eigenvalues of  $W_F$  to be less than one in absolute value. It is often used to deal with the log-determinant term that appears in the likelihood function of most spatial models (see Section 4.1). When the constraint (III) is satisfied, this log-determinant can be approximated based on a truncated Taylor series (see for example Martin 1992; Barry and Pace 1999; Pace and LeSage 2004; Smirnov and Anselin 2009; Bivand, Hauke, and Kossowski 2013). Most spatial models estimated by MLE or Bayesian methodology make use of such an approximation and thereby restrict (sometimes implicitly) the feasible parameter space according to constraint (III).

Constraints four and five can be regarded as naive higher order extensions of the previous two constraints. Their main advantage is that they do not require any complex computations. For row-stochastic neighborhood matrices, constraints (III), (IV) and (V) are equivalent if the order of the model is  $l = 1$  or when all autoregressive parameters are positive. However, the naive constraints lead to problems if we allow negative autocorrelation in a higher order spatial model. The main disadvantage of constraint (IV) is that it is much more restrictive than necessary. And constraint (V), even though it works well in most applications, does not permit to draw any conclusions about the consistency of the model.

The spatial interaction model in (6) contains three weight, matrices which is usually a blocking point for defining the feasible parameter space according to constraints (I), (II) or (III). However, the special neighborhood structure in this model allows to derive the feasible parameter space based on computations that are not much more costly than those of an order one spatial model. In model (6) the spatial filter matrix is  $W_F = \rho_d(\mathbf{I}_n \otimes W) + \rho_o(W \otimes \mathbf{I}_n) + \rho_w(W \otimes W)$  and LeSage and Pace (2008) already point out that we can compute the eigenvalues vector  $\lambda(W_F)(N \times 1)$  directly from the eigenvalues vector  $\lambda(W)(n \times 1)$  of the site neighborhood matrix  $W$ .

$$\lambda(W_F) = \rho_d(\iota_n \otimes \lambda(W)) + \rho_o(\lambda(W) \otimes \iota_n) + \rho_w(\lambda(W) \otimes \lambda(W)) \quad (9)$$

We could use (9) to verify constraint (I), but this would require up to  $O(n^3)$  computations to obtain  $\lambda(W)$  and another  $O(n^2)$  calculations for every set of values of the autoregressive parameters that we want to verify. Hence, the computational effort makes constraint (I) unusable for large sample applications.

However, if we restrict our attention to the case where  $W$  is similar to a symmetric matrix, we can verify constraints (II) or (III) without having to consider full eigenvalues vector (9) and by focusing on the relevant elements we can reduce the computational effort drastically. Let us first note that all of the  $N$  equations in (9) have the coordinate expression  $\rho_d \lambda_j(W) + \rho_o \lambda_i(W) + \rho_w \lambda_j(W) \lambda_i(W)$ , for  $i, j = 1, \dots, n$ . This coordinate form helps to see that the two extreme eigenvalues of  $W_F$  are always contained in the same four elements of the vector  $\lambda(W_F)$ : The relevant entries correspond to the case when  $\lambda_j(W)$  and  $\lambda_i(W)$  are both chosen from the set  $\{\lambda_{max}(W), \lambda_{min}(W)\}$ . This argument also applies to a general  $W$ , if none of its complex eigenvalues exceed  $\min(|\lambda_{max}(W)|, |\lambda_{min}(W)|)$  in magnitude. In the following we suppose that this condition is met <sup>3</sup> and denote the relevant entries by  $\Lambda(\rho)$ . With a small abuse of notation,

---

<sup>3</sup>It is probably possible to generalize the developments related to the definition of the feasible parameter space of model (6) to the case where  $W$  has large complex eigenvalues, but this enterprise is not the purpose of the present article. In fact, many neighborhood matrices considered in practice fulfill the requirement that none of its complex eigenvalues exceed  $\min(|\lambda_{max}(W)|, |\lambda_{min}(W)|)$  in magnitude. For example, when the neighborhood matrix is derived from contiguity or from distances it is similar to a symmetric matrix which implies that all its eigenvalues are real. Furthermore, when the neighborhood matrix  $W$  is row normalized its largest eigenvalue is real and equal to one, in which case we only have to verify that none of its complex eigenvalues exceeds  $\lambda_{min}(W)$  in magnitude.

we can define the four elements of  $\Lambda(\rho)$  by (10) below.

$$\Lambda(\rho) = \{\rho_d \lambda_j(W) + \rho_o \lambda_i(W) + \rho_w \lambda_j(W) \lambda_i(W), \quad \text{for } i, j = \min, \max\} \quad (10)$$

Verifying constraints (II) and (III) with this reduced set of relevant eigenvalues has the double advantage that we only need to compute the largest and the smallest eigenvalue<sup>4</sup> of  $W$  and that, instead of searching the minimum and the maximum of a  $(N \times 1)$  vector, we can reduce its size to a  $(4 \times 1)$  vector.

$$\begin{aligned} \text{Constraint (II)} &\iff \max(\Lambda(\rho)) < 1 \\ \text{Constraint (III)} &\iff -1 < \min(\Lambda(\rho)) \quad \& \quad \max(\Lambda(\rho)) < 1 \end{aligned} \quad (11)$$

The vector of relevant eigenvalues  $\Lambda(\rho)$  allows to verify the two constraints in (11) almost instantaneously. In the following we will use constraint (II) for the S2SLS estimator. For the MLE and the Bayesian estimator we have to use the more restrictive constraint (III) when dealing with large sample applications. In such applications, we rely on a log-determinant approximation to evaluate the likelihood function and this approximation is only valid when (III) is satisfied. To quantify the gains from choosing the spaces associated to (II) and (III) as opposed to the classical approach, we can compare their volume with the one implied by constraint (IV) requiring  $|\rho_d| + |\rho_o| + |\rho_w| < 1$ . For the example presented in Section 5, this comparison reveals that we increase the size of the feasible parameter space by a factor 2.4 when choosing (III) and by a factor 7.2 when choosing (II).

## 4 Improved estimation methods

Maximum likelihood estimation is the traditional estimation method for spatial econometric models and dates back to Ord (1975). Anselin (1988) provides an in-depth analysis of the MLE estimator for a general spatial model that may include dependence in the error term, the response variable and the explanatory variables. Much later, Lee (2004) proves asymptotic normality and consistency of the MLE and quasi-MLE and shows that the estimator attains  $\sqrt{N}$  when the neighborhood matrix is sufficiently sparse. Although Anselin (1988) already discusses alternatives to the MLE, it took another decade until Bayesian estimation (LeSage 1997) and S2SLS estimation (Kelejian and Prucha 1998) received broader attention.

All three estimators were initially proposed for a standard spatial model, but they have been adapted to the case of origin-destination flows. LeSage and Pace (2008) develop a MLE procedure for the spatial interaction model given in (6) and in their textbook (LeSage and Pace 2009, page 218) they describe an MCMC estimator of the same model. Badinger and Egger (2011) extend the S2SLS estimator of Kelejian and Prucha (1998) to models with multiple weight matrices and Tamesue and Tsutsumi (2016) use this extension for a model of migration flows in Japan. The following subsections show that we can derive all three estimators exclusively from moment matrices with low dimensions. Consequently, the computational burden of the estimation is up to the calculation of the moments independent of the number of observations. We first discuss a reformulation of the RSS term in the likelihood function that yields computational advantages for the MLE and the Bayesian MCMC estimator. Then we derive the estimators based on the moment expressions beginning with the two likelihood based ones and finally present the matrix formulation of the S2SLS estimator.

---

<sup>4</sup>Efficient algorithms that allow to extract the smallest or largest eigenvalues of large matrices are available in many software packages. Some examples are the eigs function in MATLAB or the C++ library Spectra which offers an interface to the R software.

## 4.1 Refining the likelihood expression

Computational techniques that speed up the evaluation time of the likelihood function benefit the MLE as well as Bayesian estimators. This evaluation time is largely determined by a log-determinant term and a RSS term. While the log-determinant has been studied in great detail, the RSS term is usually treated very briefly or in the context of a specific estimator. This article proceeds differently by developing the RSS term as a reformulation of the likelihood function. The RSS decomposition presented in the following draws inspiration from those of LeSage and Pace (2008) and LeSage and Pace (2009, page 50), but has the advantage that it does not require to construct additional  $(1 \times N)$  vectors of residuals and fitted values. Debarsy and LeSage (2020) derive a formulation of the RSS term that is as efficient as the one given here, but only applies to the special case of a marginal likelihood that depends on  $\rho$  but no longer on  $\delta$ . Interestingly, the concentrated log-likelihood described in Section 4.2.1 leads to the same special case as their marginal likelihood approach.

The likelihood function  $f_y(y)$  of model (6) can be derived from the density  $f_\varepsilon(\varepsilon)$  of the errors using a transformation of variables (Shao 2003, page 23) which exploits the link  $y = A^{-1}Z\delta + A^{-1}\varepsilon$  between the errors and the flows. The Jacobian term of this transformation turns out to be  $|A|$  the determinant of the spatial filter matrix.

$$f_y(y) = f_\varepsilon(\varepsilon)|A| = \left(\frac{1}{2\pi\sigma^2}\right)^{(N/2)} \exp\left\{-\frac{1}{2\sigma^2}(Ay - Z\delta)'(Ay - Z\delta)\right\} |A| \quad (12)$$

The likelihood function in (12) contains an RSS term  $RSS(\rho, \delta) = (Ay - Z\delta)'(Ay - Z\delta)$  whose evaluation would normally require  $O(n^2)$  calculations. However, we can reformulate the RSS term so that it only depends on the parameters and some low dimensional moment matrices. Let us first consider the following decomposition of  $Ay$  the filtered flow vector

$$Ay = y_\bullet \tau(\rho) \quad y_\bullet = (y^{(1)} \quad y^{(2)} \quad y^{(3)} \quad y^{(4)}) \quad \tau(\rho) = (1 \quad -\rho_d \quad -\rho_o \quad -\rho_w)', \quad (13)$$

where  $y^{(t)}$ ,  $t = 1, 2, 3, 4$  are defined in (8). The above decomposition of  $Ay$  allows to reformulate the RSS term as

$$RSS(\rho, \delta) = (Ay - Z\delta)'(Ay - Z\delta) = \tau(\rho)'TSS_\bullet\tau(\rho) - 2\delta'Z'y_\bullet\tau(\rho) + \delta'Z'Z\delta, \quad (14)$$

where  $TSS_\bullet = y'_\bullet y_\bullet$ . LeSage and Pace (2008) also make use of the  $TSS_\bullet$  matrix and point out that its elements can be computed efficiently using  $\iota'_n(Y^{(t)} \odot Y^{(s)})_{\iota_n}$ , for  $t, s = 1, 2, 3, 4$ . The MLE and the Bayesian MCMC estimators in the following two sections use the RSS expression in (14) to reduce the cost of evaluating the likelihood to a minimum.

## 4.2 The MLE

The MLE estimates of a model correspond to the parameter values that maximize the likelihood function given the data. For most spatial models these maximizers do not admit an analytical solution and we have to employ numerical optimization techniques to find them. This numerical optimization is usually costly because it requires many evaluations of the likelihood function. The following subsection recapitulates how we can reduce the dimension of this optimization problem by working with the concentrated log-likelihood function and shows how to speed up the evaluation of this function using the reformulated RSS term presented in Section 4.1. The second subsection generalizes the mixed numerical analytical Hessian approximation of LeSage and Pace (2009) to higher order spatial models and describes how to use this approximation to compute the variances of the MLE estimators without any notable effort.

### 4.2.1 The concentrated likelihood function

Applying the logarithm to the density of the dependent variable in (12) leads to the log-likelihood function which is expressed in terms of the model parameters.

$$\mathcal{L}(\rho, \delta, \sigma^2) = \ln(f_y) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (Ay - Z\delta)'(Ay - Z\delta) + \ln |A| \quad (15)$$

By differentiating the above log-likelihood with respect to the parameters  $\rho$ ,  $\delta$  and  $\sigma^2$ , we obtain the following three sets of first order conditions.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \rho} &= -\text{tr}(A^{-1} \frac{\partial A}{\partial \rho}) - \frac{(Ay - Z\delta)}{\sigma^2} \frac{\partial A}{\partial \rho} = 0 \implies \rho \text{ has no analytical expression} \\ \frac{\partial \mathcal{L}}{\partial \delta} &= -\frac{1}{\sigma^2} (Z' Ay - Z' Z\delta) = 0 \implies \hat{\delta}(\rho) = (Z' Z)^{-1} Z' Ay \\ \frac{\partial \mathcal{L}}{\partial \sigma^2} &= -\frac{N}{2\sigma^2} + \frac{(Ay - Z\delta)'(Ay - Z\delta)}{2\sigma^4} = 0 \implies \hat{\sigma}^2(\rho) = \frac{(Ay - Z\delta(\rho))'(Ay - Z\delta(\rho))}{N} \end{aligned} \quad (16)$$

The first set of equations in (16) reveals that this optimization problem does not yield an analytical expression for the autoregressive parameters. However, the solutions for the parameters  $\delta$  and  $\sigma^2$  are found explicitly as a function of  $\rho$ . We can use the notation  $y_\bullet$  in (13) to define  $\delta_\bullet = (Z' Z)^{-1} Z' y_\bullet$ , which allows to simplify the solutions to the first order conditions in (16) and also the expression of the RSS term in (14).

$$\begin{aligned} \hat{\delta}(\rho) &= \delta_\bullet \tau(\rho) \\ \hat{\sigma}^2(\rho) &= \frac{RSS(\rho, \hat{\delta}(\rho))}{N} \\ RSS(\rho, \hat{\delta}(\rho)) &= \tau(\rho)' TSS_\bullet \tau(\rho) - 2\tau(\rho)' \delta'_\bullet Z' y_\bullet \tau(\rho) + \tau(\rho)' \delta'_\bullet Z' Z \delta_\bullet \tau(\rho) \\ &= \tau(\rho)' (TSS_\bullet - (Z' y_\bullet)' (Z' Z) (Z' y_\bullet)) \tau(\rho) \\ &= \tau(\rho)' RSS_\bullet \tau(\rho) \end{aligned} \quad (17)$$

LeSage and Pace (2008) already use the factorization  $\tau(\rho)' RSS_\bullet \tau(\rho)$  of the RSS term, but they derive the  $RSS_\bullet$  matrix from an explicit expression of four residual vectors which is less efficient than deriving their inner product directly as in (17). (In vectorized form their computations correspond to  $RSS_\bullet = \varepsilon'_\bullet \varepsilon_\bullet$ , where  $\varepsilon_\bullet = y_\bullet - Z' \delta_\bullet$ .) Substituting the simplified solutions of (17) into the likelihood function  $\mathcal{L}^c(\rho) = \mathcal{L}^c(\rho, \hat{\delta}(\rho), \hat{\sigma}^2(\rho))$  yields

$$\mathcal{L}^c(\rho) = \kappa - \frac{N}{2} \ln(\tau(\rho)' RSS_\bullet \tau(\rho)) + \ln |A|, \quad (18)$$

where  $\kappa$  gathers all terms that are constant with respect to the parameters  $\rho$ . Using the log-determinant approximation presented by LeSage and Pace (2008) we can optimize this function with little computational effort. This optimization provides us with MLE estimates of  $\rho$  which are then inserted into (17) to derive the MLE estimates of  $\delta$  and  $\sigma^2$ .

### 4.2.2 The Hessian Matrix

The Hessian matrix is required to conduct inference on the parameter estimates, which is an important part of any estimation procedure. In most spatial econometric models, the expression of the Hessian matrix involves  $A^{-1}$  the inverse of the spatial filter matrix whose computational

cost are usually prohibitive. Approximations of the Hessian matrix allow to circumvent this computational bottleneck and the finite difference approximation is a very popular approach. For example, the R package `spatialreg` (Bivand and Piras 2015) or MATLAB implementations in the spatial econometrics toolbox <sup>5</sup> use it to approximate the Hessian matrix. Although the finite difference approximation avoids inverting a large matrix it requires many additional evaluations of the likelihood function, which also enhances the computational burden. The mixed numerical analytical Hessian approximation set forth in this article is effortless in comparison. This method was first introduced by LeSage and Pace (2009, page 56) who use it for a spatial model with a single weight matrix. In the following we extend their approximation method to the case of origin-destination flows with three spatial weight matrices.

The Hessian matrix corresponds to the second order derivatives of the likelihood function  $\mathcal{L}$ , as in (12), with respect to all model parameters. The variance of the error term and the parameters associated to the covariates in  $Z$  also appear in the standard regression model and they are grouped in  $\theta = (\delta' \ \sigma^2)'$ . Consequently, the likelihood function depends on two sets of parameters ( $\rho$  and  $\theta$ ) and the Hessian matrix  $H^{(b)}$  is composed of four blocks.

$$H^{(b)} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \rho \partial \rho'} & \frac{\partial^2 \mathcal{L}}{\partial \rho \partial \theta'} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \rho'} & \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta'} \end{bmatrix} = \begin{bmatrix} H_{11}^{(b)} & H_{12}^{(b)} \\ H_{21}^{(b)} & H_{22}^{(b)} \end{bmatrix} \quad (19)$$

For a spatial model of order one LeSage and Pace (2009) already demonstrate that three of the four blocks of the Hessian matrix have very simple analytical solutions. The developments in Appendix A show that their argument carries over to the case of a model with multiple weight matrices. The closed form solution of these blocks is given by

$$H_{12}^{(b)'} = H_{21}^{(b)} = \begin{bmatrix} -\frac{Z'L}{\sigma^2} \\ \frac{(Z'L)\delta' - \tau(\rho)'(y'_{\bullet}L)}{\sigma^4} \end{bmatrix} \quad H_{22}^{(b)} = \begin{bmatrix} -\frac{Z'Z}{\sigma^2} & 0 \\ 0 & -\frac{N}{2\sigma^4} \end{bmatrix}, \quad (20)$$

where the matrix  $L = (y^{(2)}y^{(3)}y^{(4)})$  groups the three spatial lags of the flow vector. The matrix  $Z'L$  corresponds to the last three columns of  $Z'y_{\bullet}$ , which are already computed for the factorization of the RSS in (17). Similarly, the matrix  $y'_{\bullet}L$  corresponds to the last three columns of the  $TSS_{\bullet}$  matrix.

The Hessian block  $H_{11}^{(b)}$  contains the second order derivatives of the likelihood function with respect to the autoregressive parameters  $\rho$ . Its analytical expression is given in (21) below, where for  $j, k \in (d, o, w)$ ,  $V_{jk} = -tr(W_j A^{-1} W_k A^{-1}) - \frac{1}{\sigma^2} y' W_j' W_k y$ .

$$H_{11}^{(b)} = \begin{bmatrix} V_{dd} & V_{do} & V_{dw} \\ \bullet & V_{oo} & V_{ow} \\ \bullet & \bullet & V_{ww} \end{bmatrix} \quad (21)$$

Computing the exact solution of  $H_{11}^{(b)}$  is infeasible in large sample applications because it depends on the inverse of the spatial filter matrix. For this reason LeSage and Pace (2009) propose to use a numerical approximation of this part of the Hessian. Their approximation makes use of a theorem (for details see Davidson and MacKinnon 1993, page 269) that links the blocks of the full Hessian matrix in (19) with the Hessian of the concentrated log-likelihood  $\mathcal{L}^c$ .

$$\frac{\partial^2 \mathcal{L}}{\partial \rho \partial \rho'} = \frac{\partial^2 \mathcal{L}^c}{\partial \rho \partial \rho'} + \frac{\partial^2 \mathcal{L}}{\partial \rho \partial \theta'} \left( \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta'} \right)^{-1} \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \rho'} \quad (22)$$

Equation (22) expresses the first Hessian block  $H_{11}^{(b)}$  in terms of the other three blocks ( $H_{12}^{(b)}$ ,  $H_{21}^{(b)}$  and  $H_{22}^{(b)}$ ) and the Hessian matrix of the concentrated likelihood  $H_{\mathcal{L}^c}$ . We can use this

<sup>5</sup><https://www.spatial-econometrics.com/>

link to derive the mixed numerical analytical approximation of  $H_{11}^{(b)}$ , which we denote  $\tilde{H}_{11}^{(b)}$ . To compute this approximation we only require the analytical estimators of the blocks given in (20) denoted by  $\hat{H}_{21}^{(b)}$  and  $\hat{H}_{22}^{(b)}$ , and a numerical estimate of the Hessian of the concentrated log-likelihood denoted by  $\tilde{H}_{\mathcal{L}^c}$ .

$$\tilde{H}_{11}^{(b)} = \tilde{H}_{\mathcal{L}^c} + \hat{H}_{21}^{(b)'} (\hat{H}_{22}^{(b)})^{-1} \hat{H}_{21}^{(b)} \quad (23)$$

The approximation of the complete Hessian then combines the mixed estimator  $\tilde{H}_{11}^{(b)}$  with the analytical estimators  $\hat{H}_{21}^{(b)}$  and  $\hat{H}_{22}^{(b)}$ . LeSage and Pace (2009) already point out that Newtonian methods for optimization will produce a numerical estimation of  $H_{\mathcal{L}^c}$  as a by-product. Consequently, if we choose such an optimization method, we can approximate the Hessian with almost no additional computational cost because all required quantities are available from the parameter estimation.

### 4.3 The Bayesian MCMC estimator

A Bayesian estimator of spatial econometric models is already mentioned by Anselin (1988), but due to its prohibitively difficult estimation it took further progress in computer technology and theory until it became a practical alternative to the MLE. The Bayesian methodology requires to assume prior distributions for all model parameters. Through the estimation we update these naive priors into posterior distributions that incorporate the information contained in the data. The MCMC methodology caused a rise in the popularity of Bayesian statistics because it avoids the main computational bottleneck of the traditional analytical approach. Instead of searching the analytical expression for the moments of the posterior distribution with numerical integration techniques, MCMC generates a random sample of parameter values. Based on this sample it is possible to infer the moments of interest, such as the expectation and variance of a parameter, using elementary statistics. Metropolis et al. (1953) solve the problem of generating random draws from the unknown posterior by constructing the sampling procedure as a Markov Chain that converges to the desired density function. Their solution is generalized by Hastings (1970) and LeSage (1997) was among the first to use a variant of the Metropolis-Hastings (M-H) algorithm to estimate a spatial econometric model. LeSage and Pace (2009) further develop this idea and estimate a spatial econometric interaction model using Bayesian MCMC methodology.

In a more recent article Fischer and LeSage (2020) develop an extension of the spatial interaction model to panel data and estimate it by Bayesian MCMC. Their estimation methodology already exploits the efficiency gains of the RSS reformulation but differs from the methods presented in this article. Similar to Debarsy and LeSage (2018) they use block-sampling in conjunction with an expression of the marginal likelihood to generate draws for the autoregressive parameters. This article, in contrast, uses a sequential sampling approach and shows that we can exploit the efficiency gains of the restructured RSS also when using the conditional instead of the marginal likelihood. Their block sampling approach has the advantage that it requires less evaluations of the likelihood function at every iteration of the MCMC sampler, but it leads to lower acceptance rates and hence requires a higher number of iterations to converge. However, since both approaches allow to express the sampling procedure in terms of low dimensional moment matrices it will be the calculation of the moments itself and not the sampling procedure that dominates the computation cost for applications to very large data sets. The following paragraphs introduce the Bayesian model of LeSage and Pace (2009) and then describe how to use the restructured RSS of Section (4.1) to increase the efficiency of the sequential sampling method.

### 4.3.1 The Bayesian model

To estimate a Bayesian model we have to assign prior distributions  $p(\phi)$  to all model parameters  $\phi$ . These prior assumption largely determine whether we can easily solve the estimation problem. The posterior distribution  $p(\phi|\mathcal{D})$  is derived from the prior assumptions and the conditional distribution of the data  $p(\mathcal{D}|\phi)$  by the Bayes rule

$$p(\phi|\mathcal{D}) = \frac{p(\mathcal{D}|\phi) \cdot p(\phi)}{p(\mathcal{D})} \propto p(\mathcal{D}|\phi) \cdot p(\phi), \quad (24)$$

where the unconditional distribution of the data  $p(\mathcal{D})$  is usually absorbed in the proportionality constant. The above expression is a standard formalization for Bayesian models and LeSage and Pace (2009) describe how to apply it to the model given in (6). In this model the set of parameters is  $\phi = (\rho', \delta', \sigma^2)'$  and the data is given by  $\mathcal{D} = (y, Z)$ . The conditional distribution of the data

$$p(\mathcal{D}|\phi) = p(\mathcal{D}|\rho, \delta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{(N/2)} \exp\left\{-\frac{1}{2\sigma^2}(Ay - Z\delta)'(Ay - Z\delta)\right\}|A| \quad (25)$$

coincides with the likelihood function used for the MLE because it is derived from the same distributional assumption on the error term  $\varepsilon \sim N(0, \sigma^2 I_N)$ .

The second term on the right hand side of (24) is defined by the prior assumptions on the model parameters, which LeSage and Pace (2009) choose carefully. Even though they point out that informative priors would only marginally influence the estimation results when the sample is sufficiently large they rely on uninformative priors for all parameters. For the autoregressive parameters  $\rho$  they choose a uniform prior which is by construction uninformative. For the parameters  $\delta$  and  $\sigma^2$  LeSage and Pace (2009) use a normal inverse gamma prior, which means that the inverse of  $\sigma^2$  follows a gamma-distribution and that  $\delta$  follows a normal distribution if  $\sigma^2$  is given. In general, these distributions are informative, but since LeSage and Pace (2009) define near infinite prior variances they avoid to introduce subjective information in the estimation. The equations in (26) below show the set of prior distributions, denoted by  $\pi$ , that the authors use to define the Bayesian model. We define the uniform prior for  $\rho$  on the parameter space  $\Theta \subset \mathbb{R}^3$  whose bounds are chosen to ensure that constraint (III) in (11) is satisfied. The resulting density is expressed using the indicator function  $\mathbb{1}_{\{x\}}$ , which is one if the event  $x$  is true and zero otherwise. LeSage and Pace (2009) note that these assumptions imply prior independence between  $\rho$  and the other parameters, which we should regard as an hypothesis that the estimation results will confirm or reject.

$$\begin{aligned} \delta|\sigma^2 \sim \mathcal{N}(c, T\sigma^2) &\implies \pi(\delta|\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{(K/2)} |T|^{(-1/2)} \exp\left\{-\frac{1}{2\sigma^2}(\delta - c)'T^{-1}(\delta - c)\right\} \\ \sigma^2 \sim IG(a, b) &\implies \pi(\sigma^2) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{\sigma^2}\right)^{(a+1)} \exp\{-b/\sigma^2\} \\ \rho \sim U(\Theta) &\implies \pi(\rho) \propto \mathbb{1}_{\{\rho \in \Theta\}} \end{aligned} \quad (26)$$

LeSage and Pace (2009) assign the value zero to  $a, b$  and  $c$  and a zero matrix to  $T^{-1}$ , which makes the distribution  $\pi(\delta|\sigma^2)$  and  $\pi(\sigma^2)$  uninformative and allows to derive the following simplified expression of the posterior density.

$$p(\delta, \sigma^2, \rho|\mathcal{D}) \propto \sigma^{(-N-1)} \exp\left\{-\frac{1}{2\sigma^2}(Ay - Z\delta)'(Ay - Z\delta)\right\}|A| \mathbb{1}_{\{\rho \in \Theta\}} \quad (27)$$

The main problem of the classical Bayesian estimator for the spatial econometric interaction model arises because the posterior in (27) is analytically not tractable. LeSage and Pace (2009)

show that the parameters  $\delta$  and  $\sigma^2$  can be integrated out to find the marginal density of the autoregressive parameters  $p(\rho|D)$ . However, it is generally not possible to find the expectation and the variance of  $p(\rho|D)$  analytically. Consequently, if we want to compute the moment expressions below we have to use numerical integration, which requires enormous computational resources.

$$E(\rho | \mathcal{D}) = \frac{\int \rho \cdot p(\rho | \mathcal{D}) d\rho}{\int p(\rho | \mathcal{D}) d\rho} \quad \text{VAR}(\rho | \mathcal{D}) = \frac{\int (\rho - E(\rho | \mathcal{D}))^2 \cdot p(\rho | \mathcal{D}) d\rho}{\int p(\rho | \mathcal{D}) d\rho}$$

### 4.3.2 The MCMC methodology

Advances in the MCMC methodology raised the popularity of Bayesian statistics because they allow to estimate a model without having to compute expressions like the above moments with numerical methods. Like ordinary Monte Carlo methods the MCMC methodology generates random draws from the distribution of the model parameters. We cannot use ordinary Monte Carlo methods because the density of the parameters is not known and MCMC solves this problem by setting up the sampling procedure as a Markov Chain that converges to the right distribution function. Since the sampler needs to be initialized with more or less arbitrary values and convergence requires some time we have to disregard the first draws as a burn-in period. When the sampler converges we can regard the remaining draws as a sample from the desired posterior distribution. LeSage and Pace (2009) explain how to set up such a MCMC sampler for the Bayesian spatial interaction model described in the previous section. They decompose the joint posterior density in (27) into three conditional densities, which leads to a multivariate normal distribution for  $\delta$ , an inverse gamma distribution for  $\sigma^2$  and an intractable density for  $\rho$ .

$$\begin{aligned} p(\delta|\sigma^2, \rho, \mathcal{D}) &= \left(\frac{1}{2\pi\sigma^2}\right)^{(K/2)} |T^*|^{(1/2)} \exp\left\{-\frac{1}{2\sigma^2}(\delta - c^*)'T^{*-1}(\delta - c^*)\right\} \sim \mathcal{N}(c^*, T^*\sigma^2) \quad (28) \\ c^* &= (Z'Z + T^{-1})^{-1}(Z' Ay + T^{-1}c) \\ T^* &= (Z'Z + T^{-1})^{-1} \end{aligned}$$

$$\begin{aligned} p(\sigma^2|\delta, \rho, \mathcal{D}) &= \frac{b^{*a^*}}{\Gamma(a^*)} \left(\frac{1}{\sigma^2}\right)^{(a^*+1)} \exp\{-b^*/\sigma^2\} \sim IG(a^*, b^*) \quad (29) \\ a^* &= a + N/2 \\ b^* &= b + (Ay - Z\delta)'(Ay - Z\delta)/2 \end{aligned}$$

$$p(\rho|\delta, \sigma^2, \mathcal{D}) \propto \left(\frac{1}{2\pi\sigma^2}\right)^{(N/2)} \exp\left\{-\frac{1}{2\sigma^2}(Ay - Z\delta)'(Ay - Z\delta)\right\} |A| \mathbb{1}_{\{\rho \in \Theta\}} \quad (30)$$

The simplifications of the densities in (29) and (30) are possible because the normal inverse gamma prior is conjugate with respect to the gaussian distribution of the errors. LeSage and Pace (2009) proceed by sampling sequentially from three conditional posteriors using a Metropolis within Gibbs sampling procedure. This procedure generates draws of  $\delta$  and  $\sigma^2$  based on Gibbs sampling and resorts to the more costly M-H algorithm only for the draws of the autoregressive parameters, which do not follow any known distribution. The next section presents the details of the sampling procedure and shows how we can use the decomposition of the RSS in Section 4.1 to make it more efficient.

### 4.3.3 The improved sampling procedure

The main issue of the MCMC sampler is that its convergence requires several thousand iterations, which amplifies the computational costs associated with the updates of each parameter. We can bring down this cost to a minimum by expressing all calculation steps in terms of moment matrices with low dimensions. These moment matrices do not change with the parameter values and hence only need to be calculated once to initiate the sampler. To distinguish these moments from quantities that change throughout the sampling procedure the following three paragraphs make extensive use of the indexation ( $i$ ) for MCMC iterations.

#### Gibbs-sampling for the parameters $\delta$

As the parameter  $\delta \sim \mathcal{N}(c^*, T^* \sigma^2)$  is conditionally normally distributed we can use Gibbs sampling for this parameter. Hence, to generate an updated draw  $\delta_{(i)}$  we only need  $c_{(i)}^*$  the updated mean of the normal distribution and  $\sigma_{(i-1)}^2$  the value of the variance parameter from the previous iteration. As LeSage and Pace (2009) we update  $\delta$  by generating a multivariate normal deviate  $\epsilon_{(i)}$  with the desired variance. This deviate is then added to the updated mean  $c_{(i)}^*$  which we compute efficiently using the simplified notation  $\delta_{\bullet}$  as in (17) and the vector  $\tau(\rho_{(i-1)}) = (1 - \rho'_{(i-1)})'$ .

$$\begin{aligned} \delta_{(i)} &= c_{(i)}^* + \epsilon_{(i)} \\ \epsilon_{(i)} &\sim \mathcal{N}(0, T^* \sigma_{(i-1)}^2) \\ c_{(i)}^* &= (Z'Z + T^{-1})^{-1}(Z'A_{(i-1)}y + T^{-1}c) = \delta_{\bullet} \tau(\rho_{(i-1)}) \\ T^* &= (Z'Z + T^{-1})^{-1} = (Z'Z)^{-1} \end{aligned} \tag{31}$$

#### Gibbs-sampling for the variance parameter $\sigma^2$

The conditional distribution of  $\sigma^2 \sim IG(a^*, b^*)$  also permits Gibbs sampling. In order to draw the updated parameter  $\sigma_{(i)}^2$  we only need to compute  $b_{(i)}^*$  which we derive from the RSS decomposition in Section 4.1.

$$\begin{aligned} \sigma_{(i)}^2 &\sim IG(a^*, b_{(i)}^*) \\ a^* &= a + \frac{N}{2} = \frac{N}{2} \\ b_{(i)}^* &= b + \frac{1}{2} (A_{(i-1)}y - Z\delta_{(i)})' (A_{(i-1)}y - Z\delta_{(i)}) = \frac{1}{2} RSS(\rho_{(i-1)}, \delta_{(i)}) \\ &= \frac{1}{2} (\tau(\rho_{(i-1)})' RSS_{\bullet} \tau(\rho_{(i-1)}) + \epsilon'_{(i)} (Z'Z) \epsilon_{(i)}) \end{aligned} \tag{32}$$

#### Metropolis-Hastings-sampling for the autoregressive parameters $\rho$

Sampling the autoregressive parameters is more complicated because the conditional posterior distribution of  $\rho$  is unknown. LeSage and Pace (2009) use a M-H sampler that incorporates the tuned random walk procedure of Holloway, Shankar, and Rahmanb (2002) to create a sample of the posterior distribution of  $\rho$ . Since this distribution is not known the M-H algorithm relies on a gaussian distribution to generate candidate draws that are accepted or rejected depending on their likelihood value.

The tuned random walk procedure ensures that the candidate values for all autoregressive parameters  $\rho_{r(i)}^c$ ,  $r = d, o, w$  are accepted with a probability between 40% and 60%. This is necessary because too high or too low acceptance rates indicate that generated parameter draws

do not represent the entire posterior distribution. The acceptance rate of the candidates is directly linked to the variance of the gaussian proposal distribution and the optimal choice for this variance depends on the sample size. To make the estimator more flexible this variance is controlled by the tuning parameter  $c_r$ , as the below equation illustrates.

$$\rho_{r(i)}^c = \rho_{r(i-1)} + c_r \cdot \mathcal{N}(0, 1) \quad \text{for } r = d, o, w$$

The tuning parameter is adjusted whenever the acceptance rate of the proposal distribution lies outside the desired interval. In case of a too low acceptance rate the parameter  $c_r$  is divided by 1.1, and in case of a too high acceptance rate it is multiplied by 1.1. LeSage and Pace (2009) also point out that restricting the candidates to the feasible parameter space can be done by discarding the candidates until the chosen validity constraints are met.

Given a valid candidate  $\rho_{r(i)}^c$  (for  $r = d, o, w$ ) we accept it as an actual draw from the conditional posterior distribution if it leads to a higher likelihood value than the previous draw of the same parameter  $\rho_{r(i-1)}$ . When the likelihood value is lower we still accept it with a probability that is computed as the ratio of the conditional posteriors evaluated at the candidate and the previous value. If a new candidate is accepted the current value is set to the candidate value  $\rho_{r(i)} = \rho_{r(i)}^c$ , otherwise it remains at the previous value  $\rho_{r(i)} = \rho_{r(i-1)}$ . We illustrate this calculation only for the draws of  $\rho_d$ , because updating the draws for the other autoregressive parameters ( $\rho_o$  and  $\rho_w$ ) can be done with almost identical calculations. While the order of updates is irrelevant, it is important to consider the updated values of all parameters when updating the next one. To evaluate the candidate value  $\rho_{d(i)}^c$  we first have to compute its acceptance probability according to the equations below.

$$P(\rho_{d(i)} = \rho_{d(i)}^c) = 1 - P(\rho_{d(i)} = \rho_{d(i-1)}) = \min \left\{ 1, LR(\rho_{d(i)}^c) \right\} \quad (33)$$

$$LR(\rho_{d(i)}^c) = \frac{p(\rho_{d(i)}^c | \delta_{(i)}, \sigma_{(i)}^2, \rho_{o(i-1)}, \rho_{w(i-1)}, \mathcal{D})}{p(\rho_{d(i-1)} | \delta_{(i)}, \sigma_{(i)}^2, \rho_{o(i-1)}, \rho_{w(i-1)}, \mathcal{D})}$$

The largest part of the computational cost of the estimation procedure is linked to the evaluation of the above likelihood ratio and it is more convenient to evaluate its logarithm. This requires to compute the difference between two log-determinant terms and the difference between two RSS terms which we denote  $\Delta RSS(\rho_{d(i)}^c)$ .

$$\ln LR(\rho_{d(i)}^c) = \ln |A(\rho_{d(i)}^c, \rho_{o(i-1)}, \rho_{w(i-1)})| - \ln |A(\rho_{d(i-1)})| - \frac{\Delta RSS(\rho_{d(i)}^c)}{2\sigma_{(i)}^2} \quad (34)$$

We can compute the difference in the RSS terms almost instantaneously using

$$\Delta RSS(\rho_{d(i)}^c) = \tau_{d(i)}^{c'} RSS_{\bullet} \tau_{d(i)}^c - \tau_{d(i-1)}' RSS_{\bullet} \tau_{d(i-1)}, \quad (35)$$

where  $\tau_{d(i)}^c = (1 - \rho_{d(i)}^c - \rho_{o(i-1)} - \rho_{w(i-1)})$ . Expression (35) can be derived from the reformulation of the RSS term, as is already demonstrated for the computation of  $b_{(i)}^*$  in (32). Based on these developments we can compute the acceptance probability in (33) at a cost that is not much higher than the cost a single log-determinant approximation, as the second log-determinant is always available from the previous parameter update.

#### 4.4 The S2SLS estimator

The spatial two-stage least-squares estimator is essentially an instrumental variable estimation that treats the spatial lags of the dependent variables as endogenous regressors. This framing

of the estimation problem is already developed by Anselin (1988), but the S2SLS became an operational estimation method only after Kelejian and Prucha (1998) demonstrate that suitable instrumental variables can be derived from higher order spatial lags of the exogenous covariates. The authors prove consistency and asymptotic normality of their two-stage estimation and subsequently extend their methodology to more complex models (see for example Kelejian and Prucha 2004; Kelejian and Prucha 2010). Badinger and Egger (2011) extend the S2SLS estimator to spatial models with multiple weight matrices and Tamesue and Tsutsumi (2016) use this extension of the S2SLS to estimate a model of migration flows in Japan. However, as the authors focus on the issue of missing values for intra-regional flows, they do not discuss potential multicollinearity problems of the instrumental variables. Such problems arise systematically when the S2SLS estimator is used for a spatial interaction models and the following paragraphs show how they can be avoided. Additionally, we show how adapt the two-stages of the estimation so that we can exploit the efficiency gains provided by the matrix formulation.

#### 4.4.1 Instrumental variables for interaction models

When estimating a spatial interaction model with S2SLS we consider the spatial lags of the dependent variable  $L = (W_d y \ W_o y \ W_w y)$  as endogenous regressors in an ordinary linear model which leads to the following reformulation of model (6).

$$y = L\rho + Z\delta + \varepsilon \quad (36)$$

We can solve the endogeneity issue by using instrumental variables for the endogenous regressors and Kelejian and Prucha (1998) show that suitable instruments for this type of model can be easily found among higher order spatial lags of the covariates in the matrix  $Z$ . While the notion of higher order spatial lags is straightforward for models with a single weight matrix, it becomes a bit more complicated in higher order models. Badinger and Egger (2011) explain how to generate such lags for models with an arbitrary number of spatial weight matrices. For the model in (36) we derive the required lag-operators from the interactions

$$\mathbf{I}_N^{\omega_1} \times W_d^{\omega_2} \times W_o^{\omega_3} \times W_w^{\omega_4}, \quad (37)$$

where the order of lags is given by  $m$ . The powers associated with the weight matrices are given by  $\omega_t = 1, \dots, m$  (for  $t = 1, 2, 3, 4$ ) and their sum  $\sum_{t=1}^4 \omega_t \leq m$  should not exceed the defined order of spatial lags. For their model of Japanese migration flows Tamesue and Tsutsumi (2016) set  $m = 2$ , which is the common choice for the order of lags that is used to generate instruments for the S2SLS estimator (Kelejian and Prucha 1998; Badinger and Egger 2011). Tamesue and Tsutsumi (2016) rightfully note that many of the lag operators generated by the interactions in (37) are in fact redundant and they reduce them to the set of unique lag operators which we define as  $W_U^m$ . For  $m = 2$  this set is shown in (38) below.

$$W_U^2 = \{\mathbf{I}_N \ W_d \ W_o \ W_d^2 \ W_o^2 \ W_w \ W_w^2 \ W_d W_w \ W_o W_w\} \quad (38)$$

However, even if we only use this reduced set of lag operators many of the instrument that we derive from  $Z$  will be duplicated which leads to collinearity problems during the estimation. For this reason, we associate specific lag operators in  $W_U^2$  to each of the six sets of variables contained in the matrix  $Z$  (see equations (5)). The resulting instrumental variables are grouped into four sets  $U = (U_\alpha \ U_{\alpha_I} \ U_\beta \ U_\gamma)$ , which are described in the sequel.

The first set of instruments  $U_\alpha$  corresponds to the global constant. We derive no instruments from this variable because constants are invariant to multiplication with row stochastic matrices.

$$U_\alpha = \iota_n \otimes \iota_n = \iota_N \quad (39)$$

The instruments in  $U_{\alpha_I}$  are derived from the intra-regional constant  $\iota_I = \text{VEC}(\mathbf{I}_n)$  which is multiplied by all the lag operators in  $W_U^2$ .

$$U_{\alpha_I} = \{\mathbf{I}_N \quad W_d \quad W_o \quad W_d^2 \quad W_o^2 \quad W_w \quad W_w^2 \quad W_d W_w \quad W_o W_w\} \times \iota_I \quad (40)$$

The resulting variables can be derived from the site neighborhood matrix  $W(n \times n)$  and do not require to actually construct any of the  $(N \times N)$  lag operators contained in  $W_U^2$ .

$$\begin{array}{lll} \mathbf{I}_N \iota_I = \text{VEC}(\mathbf{I}_n) & W_d^2 \iota_I = \text{VEC}(W^2) & W_w^2 \iota_I = \text{VEC}(W^2 W^{2'}) \\ W_d \iota_I = \text{VEC}(W) & W_o^2 \iota_I = \text{VEC}(W^{2'}) & W_d W_w \iota_I = \text{VEC}(W^2 W') \\ W_o \iota_I = \text{VEC}(W') & W_w \iota_I = \text{VEC}(W W') & W_o W_w \iota_I = \text{VEC}(W W^{2'}) \end{array}$$

The matrix  $U_\beta$  contains all instruments derived from the origin and destination attributes. When constructing these instruments, we have to pay attention to two mechanisms which, if we ignore them, create duplicated columns in the instrument matrix. One of these mechanisms results from the row-stochastic property of the  $W$  matrix, a property that makes it unnecessary to apply the origin lag operator  $W_o$  to the destination attributes  $\tilde{X}_d$  (as  $W_o^m \tilde{X}_d = \tilde{X}_d$  for  $m = 0, 1, 2, \dots$ ). The same argument holds when applying powers of the destination lag operator  $W_d$  to the origin attributes  $\tilde{X}_o$ . For the same reason it is also superfluous to apply origin-to-destination lag operator  $W_w$  to the site attributes (as  $W_w^m \tilde{X}_d = W_d^m \tilde{X}_d$  and  $W_w^m \tilde{X}_o = W_o^m \tilde{X}_o$  for  $m = 0, 1, 2, \dots$ ). The second mechanism causes duplication in the SDM specification, where we use a spatial lag of the covariates as additional explanatory variables. In this case, instruments that are generated as  $W_d^m \tilde{X}_d$  and  $W_o^m \tilde{X}_o$  will duplicate the first  $m - 1$  lags.

We can avoid duplication of instrumental variables altogether if we define them according to the equations in (41) below. These instruments are derived from lags of the site attributes ( $DX, OX$  and  $IX$ ), where the order of lags is given by  $\tilde{m} = m + 1$  which is incremented by one if the variables are used according to the SDM specification. Given the lagged site attributes we use Kronecker products to derive the instruments  $U_\beta = (\tilde{X}_d \tilde{X}_o \tilde{X}_I)$ .

$$\begin{array}{ll} \tilde{DX} = (W^0 DX \quad W^1 DX \quad \dots \quad W^{\tilde{m}} DX) & \tilde{X}_d = \iota_n \otimes \tilde{DX} \\ \tilde{OX} = (W^0 OX \quad W^1 OX \quad \dots \quad W^{\tilde{m}} OX) & \tilde{X}_o = \tilde{OX} \otimes \iota_n \\ \tilde{IX} = (W^0 IX \quad W^1 IX \quad \dots \quad W^{\tilde{m}} IX) & \tilde{X}_I = \text{Diag}(\iota_I)(\iota_n \otimes \tilde{IX}) \end{array} \quad (41)$$

The fourth set of instruments  $U_\gamma$  is derived from the attributes of the origin-destination pairs, denoted by  $g$ . Applying all the lag operators in (38) to this variable does not generally lead to duplicated instruments. However, using all these lags is not necessary and increases the computational burden of the whole estimation procedure significantly. For this reason, we will only use powers of the origin-to-destination lag operator  $W_w$  to derive instruments from  $g$ . The choice of  $W_w$  over  $W_o$  and  $W_d$  is more appealing because attributes such as the distance are only defined for origin-destination pairs, but not for the sites. We can then compute  $\check{G} = W G W'$  and  $\check{\check{G}} = W \check{G} W'$ , using  $G(n \times n)$  the matrix representation of the distances and the neighborhood matrix  $W(n \times n)$  of the sites. Based on the lags of  $G$  we can derive the instruments in  $U_\gamma$ .

$$U_\gamma = \{\mathbf{I}_N \quad W_w \quad W_w^2\} \times g = \left( \text{VEC}(G) \quad \text{VEC}(\check{G}) \quad \text{VEC}(\check{\check{G}}) \right), \quad (42)$$

#### 4.4.2 Moment based S2SLS estimation

Given the previously defined instruments we can carry out the S2SLS estimation to identify the parameter values of model (36). The first step of the estimation procedure is given by

$$\hat{L} = P_U L = U(U'U)^{-1}U'L, \quad (43)$$

where  $P_U$  denotes the projection matrix. This step consists of projecting the endogenous regressors on the space spanned by the columns of the instrument matrix. The second stage uses the fitted values of the first stage  $\hat{L}$  as additional covariates and estimates the model parameters  $\mu' = (\rho'\delta')$  using an OLS regression of  $y$  on  $\hat{Z} = (\hat{L} \ Z)$ .

$$\hat{\mu} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'y \quad (44)$$

Even though the above two stages are simple to understand and implement, they involve computations with the matrices  $U$ ,  $Z$  and  $\hat{L}$  which are large for applications with many origin-destination pairs. In the following we will derive an estimation procedure that circumvents the construction of these three matrices. Based on this procedure we can derive the S2SLS estimates only from moment expressions whose dimensions are independent of the sample size. Equation (44) shows that estimating the model parameters only requires the two moments  $(\hat{Z}'\hat{Z})$  and  $\hat{Z}'y$  which are defined in terms of the five blocks shown below.

$$\hat{Z}'\hat{Z} = \begin{pmatrix} \hat{L}'\hat{L} & \hat{L}'Z \\ \bullet & Z'Z \end{pmatrix} \quad \hat{Z}'y = \begin{pmatrix} \hat{L}'y \\ Z'y \end{pmatrix} \quad (45)$$

Two of the five blocks in (45) coincide with the moment matrices  $Z'Z$  and  $Z'y$ , defined by (7) and (8). The first block of the  $\hat{Z}'\hat{Z}$  moment corresponds to the inner product of the fitted values of the first stage in (43). We can derive this inner product directly from the two moments  $U'U$  and  $U'L$  without having to compute  $\hat{L}$  ( $N \times 3$ ) explicitly.

$$\hat{L}'\hat{L} = L'P'_U P_U L = L'P_U L = (U'L)'(U'U)^{-1}(U'L) \quad (46)$$

To compute the  $\hat{L}'Z$  block in the  $\hat{Z}'\hat{Z}$  moment we also do not require the fitted values of the first stage. In fact we can omit the projection  $P_U$  entirely because the matrix  $Z$  is a subset of the columns of the matrix  $U$ .

$$\hat{L}'Z = L'P'_U Z = L'(P_U Z) = L'Z \quad (47)$$

The last block required to define the second stage moments in (45) is  $\hat{L}'y$  which can be done efficiently using the following developments.

$$\hat{L}'y = L'P_U y = (U'L)'(U'U)^{-1}(U'y) \quad (48)$$

The equations (48), (47) and (46) show that it is possible to derive the matrices required for the second stage of the estimation procedure directly from the three moments  $U'U$ ,  $U'L$  and  $U'y$ . Since  $Z'Z$  and  $Z'y$  correspond to subsets of the columns and rows of these three moments they do not add to the computational burden of the estimation procedure.

Hence, to make the S2SLS estimation as efficient as possible we have to derive the three moments  $U'U$ ,  $U'L$  and  $U'y$  from the matrix formulation of the model. The columns of  $U'L$  and  $U'y$  have identical structures and we can use the definition of  $y^{(t)}$  ( $t = 1, 2, 3, 4$ ) in (13) to compute them efficiently. The block-structure of the required moments is derived from the four instrument groups  $U_\alpha$ ,  $U_{\alpha_I}$ ,  $U_\beta$  and  $U_\gamma$ .

$$U'U = \begin{pmatrix} U_{\alpha\alpha} & U_{\alpha\alpha_I} & U_{\alpha\beta} & U_{\alpha\gamma} \\ \bullet & U_{\alpha_I\alpha_I} & U_{\alpha_I\beta} & U_{\alpha_I\gamma} \\ \bullet & \bullet & U_{\beta\beta} & U_{\beta\gamma} \\ \bullet & \bullet & \bullet & U_{\gamma\gamma} \end{pmatrix} \quad U'y^{(t)} = \begin{pmatrix} U'_\alpha y^{(t)} \\ U'_{\alpha_I} y^{(t)} \\ U'_\beta y^{(t)} \\ U'_\gamma y^{(t)} \end{pmatrix} \quad (49)$$

Detailed calculations for all 14 blocks of the above moments in can be found in the Appendix B. Given these moments we can derive the two matrices in (45) efficiently and continue to estimate  $\hat{\mu} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'y$ . We can then compute  $\hat{\sigma}^2 = y'y - (\hat{Z}'y)'\hat{\mu}$  which allows to derive the variance-covariance matrix of  $\hat{\mu}$ .

$$\widehat{Var}(\hat{\mu}) = \hat{\sigma}^2(\hat{Z}'\hat{Z})^{-1},$$

## 5 Applied example of commuting flows

All three estimators are used for a model of home-to-work commuting flows between the municipalities around the center of Paris. The example includes  $n = 71$  regions that are characterized by their population, the median income and the number of companies located in the region.

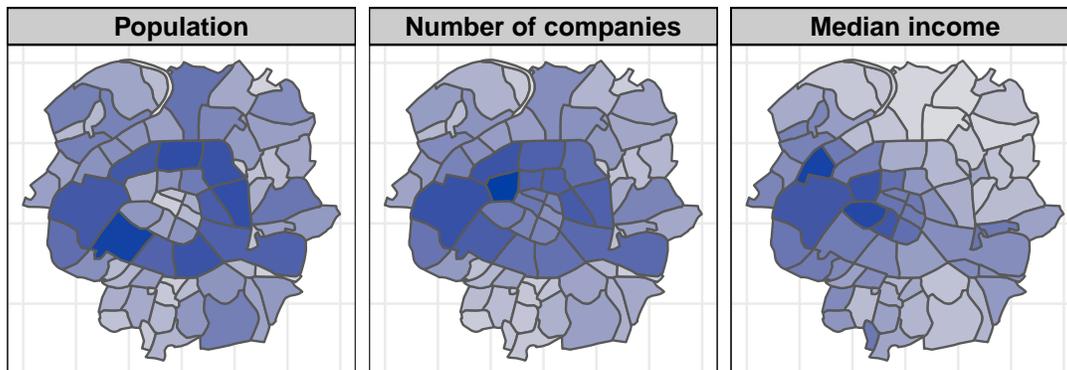


Figure 1: Explanatory variables for the municipalities of Paris (values are transformed to logarithms).

Figure 1 illustrates the spatial distribution of the region-attributes, which already show signs of spatial dependence. We then use a subset of these covariates as origin, destination, and intra-regional characteristics and also the pairwise distances to estimate three different models that explain the commuting flows between the municipalities. Two of these models make use of spatially lagged variables and for both spatial models we suppose that the neighborhood structure is defined by the contiguity of the municipalities. The most exhaustive model corresponds to a SDM and includes the three spatial autoregressive components shown in (6) and some spatial lags of the exogenous variables. This SDM is estimated by the three estimation methods of Section 4. In addition to the SDM we estimate a model that uses the same covariates but does not account for spatial autocorrelation in the flows. Since this model contains some spatial lags among its exogenous covariates it is called spatial linear model (SLM) in the sequel. The third model is an ordinary linear model (OLM) that does not include any spatially lagged variable. The ordinary as well as the spatial linear model are estimated by OLS.

The resulting parameter estimates and their standard errors are presented in Table 2. For MLE and Bayesian MCMC the results are almost identical, while the S2SLS yields parameter estimates that have the same sign as those of the likelihood based estimators but quite different magnitudes. This pattern also holds for the OLS results of the SLM and the OLM, with an

Table 2: Estimation results

Parameter	MLE (SDM)	MCMC (SDM)	S2SLS (SDM)	OLS (SLM)	OLS (OLM)
$\rho_d$	0.268* (0.018)	0.269* (0.019)	0.135 (0.102)		
$\rho_o$	0.674** (0.011)	0.673** (0.01)	0.665* (0.099)		
$\rho_w$	-0.138' (0.022)	-0.138' (0.023)	0.069 (0.086)		
<b>Constant</b>	2.651* (0.251)	2.661* (0.248)	1.89 (0.643)	15.166** (0.177)	14.958** (0.177)
<b>Constant (intra)</b>	1.267* (0.187)	1.258* (0.188)	1.609' (0.44)	-6.498** (0.199)	-6.292* (0.199)
<b>Dest: clog(MED_INCOME + 1)</b>	0.345* (0.015)	0.346* (0.015)	0.347' (0.107)	1.053** (0.014)	0.978** (0.011)
<b>Dest: clog(NB_COMPANY + 1)</b>	0.008 (0.032)	0.009 (0.033)	-0.423' (0.134)	0.25' (0.046)	0.069 (0.042)
<b>Dest: clog(MED_INCOME + 1).lag</b>	-0.179* (0.026)	-0.18* (0.026)	-0.251 (0.103)	-0.303* (0.032)	
<b>Dest: clog(NB_COMPANY + 1).lag</b>	0.073 (0.034)	0.075 (0.034)	0.651' (0.184)	0.373* (0.05)	
<b>Orig: clog(MED_INCOME + 1)</b>	0.114' (0.02)	0.114' (0.02)	-0.061 (0.034)	-0.292* (0.026)	-0.302* (0.02)
<b>Orig: clog(NB_COMPANY + 1)</b>	-0.059 (0.052)	-0.059 (0.052)	-0.031 (0.042)	-0.051 (0.077)	-0.108 (0.046)
<b>Orig: clog(POPULATION + 1)</b>	0.589* (0.026)	0.589* (0.026)	0.903* (0.111)	1.261** (0.029)	1.266** (0.027)
<b>Orig: clog(MED_INCOME + 1).lag</b>	-0.319* (0.02)	-0.318* (0.021)	-0.01 (0.045)	-0.014 (0.023)	
<b>Orig: clog(POPULATION + 1).lag</b>	0.077 (0.069)	0.077 (0.07)	-0.654' (0.122)	-0.079 (0.103)	
<b>Intra: clog(POPULATION + 1)</b>	-0.467' (0.084)	-0.469' (0.085)	-0.443' (0.113)	-0.825* (0.124)	-0.769' (0.125)
<b>log(DISTANCE + 1)</b>	-0.208* (0.022)	-0.209* (0.022)	-0.153 (0.054)	-1.213** (0.02)	-1.19** (0.02)
<b>N</b>	5041	5041	5041	5041	5041
<b>R<sub>corr</sub><sup>2</sup></b>	0.917	0.917	0.919	0.816	0.812
<b>Time (ms)</b>	79.906	1749.422	222.339	44.512	38.205

*Notes:*

Standard errors are given in parenthesis.

The function clog(x) centers the logarithm of x.

The execution time is calculated as a median over 15 repetitions.

The significance codes are: 0.1%(\*\*\*), 1%(\*\*), 5%(\*), 10%(')

exception for the parameter of the intra-regional constant. The first three rows of Table 2 report the results for the autoregressive parameters in the SDM. The parameter  $\rho_d$  of destination

dependence is significant and positive for MLE and Bayesian MCMC, but insignificant for S2SLS estimation. All three estimators of the SDM model report a significant and positive origin dependence parameter  $\rho_o$  and an insignificant origin-to-destination dependence parameter  $\rho_w$ .

To compare the goodness of fit of the five estimations we use the  $R^2_{\text{corr}}$  measure which is computed as the squared correlation between the actual flows and their fitted values<sup>6</sup>. For the two models that are estimated by OLS this measure coincides with the usual coefficient of determination ( $R^2$ ) and reveals that the OLM explains 81.2% and the SLM explains 81.6% of the variation in the data. For all three estimates of the SDM the  $R^2_{\text{corr}}$  is just below 92%. In terms of the computation time<sup>7</sup> the OLS is, not surprisingly, the fastest of the four estimation methods, but the MLE and S2SLS are also computed within fractions of a second. For this example the Bayesian MCMC is, with about 1.75 seconds, the slowest method, which is explained by the 5500 iterations that are used to generate the MCMC sample.

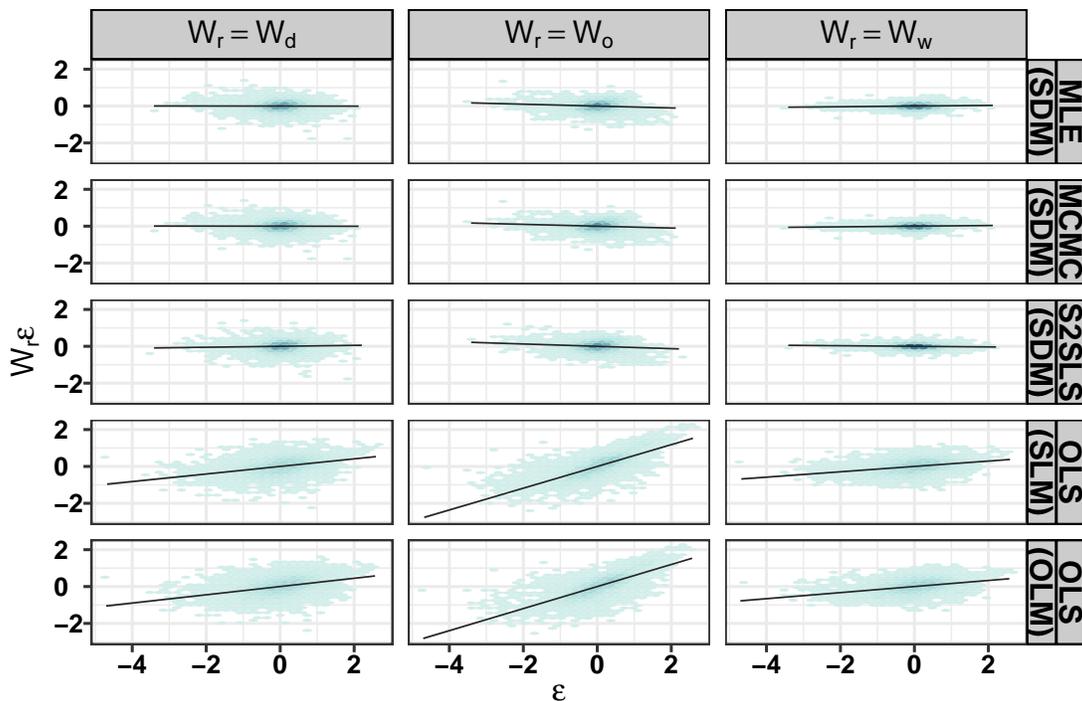


Figure 2: Moran scatter plot of residuals against spatial lags of residuals. To avoid overplotting the point clouds are converted into bi-variate densities, where darker colors indicate a high density of points.

Figure 2 shows a matrix of bi-variate scatter plots of the residuals against their spatial lags. The rows correspond to the five estimations, while the columns represent the three neighborhood matrices that are used to compute the lagged versions of the residuals  $W_r \cdot \varepsilon$ , for  $r = d, o, w$ . The slopes of the regression lines in the first three rows and columns of Figure 2 indicate that the residuals of the SDM are not spatially correlated for all estimators presented in Section 4.

<sup>6</sup>The fitted values are computed according to the in-sample trend signal (TS) predictor which is described by Goulard, Laurent, and Thomas-Agnan (2017)

<sup>7</sup>All computations are performed on the same system; Ubuntu (20.04.1 LTS), R version 4.0.3, Intel Core i5-7200U CPU with a base frequency of 2.50GHz.

However, for the OLM and SLM models there is noticeable positive spatial autocorrelation for all three definitions of the spatial neighborhood (see rows four and five of Figure 2). The strongest spatial dependence is observed for the origin neighborhood  $W_o$  displayed in the second column of Figure 2. Destination dependence and origin-to-destination dependence in the residuals of the SLM and the OLM are weaker magnitude but still clearly different from zero.

## 6 Conclusion

This article develops improved calculation techniques alongside a generalized version of the matrix formulation for spatial econometric interaction models that reduce the computational burden of the estimation considerably and allow to use the covariates more flexibly. Additionally, this article shows that the spatial econometric interaction model does not suffer from issues related to the definition of the feasible parameter space that typically arise in higher order spatial models. Thanks to a reformulation of the likelihood function and an extension of the mixed numerical analytical Hessian to spatial models with multiple weight matrices the computational requirement of our MLE estimator becomes comparable to the requirements of an OLS estimation of an ordinary linear model. The superiority of the spatial model is again confirmed by the results of our application to home-to-work commuting flows. Irrespective of whether the model is estimated by MLE, Bayesian MCMC or S2SLS, we find significant autocorrelation parameters, indicating that a model which ignores spatial autocorrelation is misspecified. Moreover, the spatial estimates lead to a better goodness of fit and yield regression residuals that are in line with the assumptions of independent errors that is used to define the model. Consequently, the spatial econometric model presents itself as the better alternative to the traditional gravity model and should be preferred whenever possible. The improvements presented by this article show that spatial econometric interaction models can be used for very large data sets, but further generalizations are necessary before they can be considered as a drop-in replacement for the traditional gravity model. Extensions that handle a sparse flow matrix or the case where the list of origins is different from the list destinations are interesting paths for future research.

## Acknowledgment

I acknowledge funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d’Avenir) program, grant ANR-17-EURE-0010, from the French National Association of Research and Technology (ANRT) and from the market research agency BVA.

## References

- Anselin, Luc (1988). *Spatial Econometrics: Methods and Models*. Springer Netherlands.
- Badinger, Harald and Peter Egger (2011). “Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances”. In: *Papers in Regional Science*, pp. 213–235.
- Barry, Ronald Paul and Robert Kelley Pace (1999). “Monte Carlo estimates of the log determinant of large sparse matrices”. In: *Linear Algebra and its Applications*, pp. 41–54.
- Bivand, Roger, Jan Hauke, and Tomasz Kossowski (2013). “Computing the Jacobian in Gaussian Spatial Autoregressive Models: An Illustrated Comparison of Available Methods”. In: *Geographical Analysis*, pp. 150–179.

- Bivand, Roger and Gianfranco Piras (2015). “Comparing Implementations of Estimation Methods for Spatial Econometrics”. In: *Journal of Statistical Software*, pp. 1–36.
- Curry, Leslie (1972). “A spatial analysis of gravity flows”. In: *Regional Studies*, pp. 131–147.
- Davidson, Russell and James G. MacKinnon (1993). *Estimation and Inference in Econometrics*. Oxford University Press.
- Debarsy, Nicolas and James Paul LeSage (2018). “Flexible dependence modeling using convex combinations of different types of connectivity structures”. In: *Regional Science and Urban Economics*, pp. 48–68.
- (2020). “Bayesian Model Averaging for Spatial Autoregressive Models Based on Convex Combinations of Different Types of Connectivity Matrices”. In: *Journal of Business & Economic Statistics*, pp. 1–12.
- Elhorst, J. Paul (2010). “Applied Spatial Econometrics: Raising the Bar”. In: *Spatial Economic Analysis*, pp. 9–28.
- Elhorst, J. Paul, Donald J. Lacombe, and Gianfranco Piras (2012). “On model specification and parameter space definitions in higher order spatial econometric models”. In: *Regional Science and Urban Economics*, pp. 211–220.
- Fischer, Manfred M. and James Paul LeSage (2020). “Network dependence in multi-indexed data on international trade flows”. In: *Journal of Spatial Econometrics*, p. 4.
- Goulard, Michel, Thibault Laurent, and Christine Thomas-Agnan (2017). “About predictions in spatial autoregressive models: optimal and almost optimal strategies”. In: *Spatial Economic Analysis*, pp. 304–325.
- Hastings, W. K. (1970). “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”. In: *Biometrika*, pp. 97–109.
- Hepple, L W (1995). “Bayesian Techniques in Spatial and Network Econometrics: 2. Computational Methods and Algorithms”. In: *Environment and Planning A: Economy and Space*, pp. 615–644.
- Holloway, Garth, Bhavani Shankar, and Sanzidur Rahmanb (2002). “Bayesian spatial probit estimation: a primer and an application to HYV rice adoption”. In: *Agricultural Economics*, pp. 383–402.
- Kelejian, Harry H. and Ingmar R. Prucha (1998). “A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances”. In: *The Journal of Real Estate Finance and Economics*, pp. 99–121.
- (2004). “Estimation of simultaneous systems of spatially interrelated cross sectional equations”. In: *Journal of Econometrics*, pp. 27–50.
- (2010). “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances”. In: *Journal of Econometrics*, pp. 53–67.
- Kelejian, Harry H. and Dennis P. Robinson (1995). “Spatial Correlation: A Suggested Alternative to the Autoregressive Model”. In: ed. by Luc Anselin and Raymond J. G. M. Florax. Springer, pp. 75–95.
- Kerkman, Kasper, Karel Martens, and Henk Meurs (2017). “A multilevel spatial interaction model of transit flows incorporating spatial and network autocorrelation”. In: *Journal of Transport Geography*, pp. 155–166.
- Lee, Lung-Fei (2004). “Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models”. In: *Econometrica*, pp. 1899–1925.
- Lee, Ming-Long and Robert Kelley Pace (2005). “Spatial Distribution of Retail Sales”. In: *The Journal of Real Estate Finance and Economics*, pp. 53–69.
- LeSage, James Paul (1997). “Bayesian Estimation of Spatial Autoregressive Models”. In: *International Regional Science Review*, pp. 113–129.
- LeSage, James Paul and Robert Kelley Pace (2008). “Spatial Econometric Modeling of Origin-Destination Flows”. In: *Journal of Regional Science*, pp. 941–967.

- LeSage, James Paul and Robert Kelley Pace (2009). *Introduction to Spatial Econometrics*. CRC Press.
- Martin, R. J. (1992). “Approximations to the determinant term in gaussian maximum likelihood estimation of some spatial models”. In: *Communications in Statistics - Theory and Methods*, pp. 189–205.
- Metropolis, Nicholas et al. (1953). “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics*, pp. 1087–1092.
- Ord, Keith (1975). “Estimation Methods for Models of Spatial Interaction”. In: *Journal of the American Statistical Association*, pp. 120–126.
- Oshan, Taylor M (2020). “The spatial structure debate in spatial interaction modeling: 50 years on”. In: *Progress in Human Geography*, p. 0309132520968134.
- Pace, Robert Kelley and James Paul LeSage (2004). “Chebyshev approximation of log-determinants of spatial weight matrices”. In: *Computational Statistics & Data Analysis*, pp. 179–196.
- Porojan, A. (2001). “Trade Flows and Spatial Effects: The Gravity Model Revisited”. In: *Open Economies Review*, pp. 265–280.
- Shao, Jun (2003). *Mathematical Statistics*. Springer New York.
- Smirnov, Oleg and Luc Anselin (2009). “An O(N) parallel method of computing the Log-Jacobian of the variable transformation for models with spatial interaction on a lattice”. In: *Computational Statistics & Data Analysis*, pp. 2980–2988.
- Tamesue, Kazuki and Morito Tsutsumi (2016). “Dealing with Intraregional Flows in Spatial Econometric Gravity Models”. In: *Spatial Econometric Interaction Modelling*. Ed. by Roberto Patuelli and Giuseppe Arbia. Springer International Publishing. Chap. 6, pp. 105–119.
- Tiefelsdorf, M. (2003). “Misspecifications in interaction model distance decay relations: A spatial structure effect”. In: *Journal of Geographical Systems*, pp. 25–50.
- Wilson, A. G. (1967). “A statistical theory of spatial distribution models”. In: *Transportation Research*, pp. 253–269.
- Young, Ernest Charles (1924). *The Movement of Farm Population*. Cornell University Agricultural Experiment Station.

## Appendix A Hessian matrix

This appendix develops the simplified analytical expression of the elements of the Hessian matrix that are presented in Section 4.2.2. These expressions are derived from the second order derivatives of the likelihood function given in (15). LeSage and Pace (2009) derive similar expressions for a spatial model of order one. Another treatment of the Hessian matrix for the spatial econometric models can be found in Anselin (1988, see pages 74ff.), who provides many computation steps linked to derivatives of the spatial filter matrix that are similar to the ones shown below.

$$\begin{aligned} \frac{\partial A}{\partial \rho_d} &= -W_d & \frac{\partial \ln|A|}{\partial \rho_d} &= \text{tr}(A^{-1} \frac{\partial A}{\partial \rho_d}) = -\text{tr}(A^{-1}W_d) \\ \frac{\partial A'A}{\partial \rho_d} &= -2A'W_d & \frac{\partial A^{-1}}{\partial \rho_d} &= -A^{-1} \frac{\partial A}{\partial \rho_d} A^{-1} = A^{-1}W_d A^{-1} \end{aligned}$$

The following paragraphs show that the Hessian matrix of higher order spatial models, in particular the one of model (6), also admits very simple analytical solutions to most of its elements. Since the expression of the likelihood function is invariant by permutation of the terms  $\rho_o W_o$ ,  $\rho_d W_d$ ,  $\rho_w W_w$  many elements of the Hessian matrix are up to the indexes ( $o, d, w$ ) identical. This property allows shorten the following argumentation, since we only have to develop the expressions related to  $\rho_d$ , while the expressions related to  $\rho_o$  and  $\rho_w$  are inferred by exchanging the indexes.

### Block $H_{11}^{(b)}$

The Hessian block  $H_{11}^{(b)}$  given in (21) is the second order derivative of the likelihood function with respect to  $\rho$

$$H_{11}^{(b)} = \frac{\partial^2 \mathcal{L}}{\partial \rho \partial \rho'} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \rho_d^2} & \frac{\partial^2 \mathcal{L}}{\partial \rho_d \partial \rho_o} & \frac{\partial^2 \mathcal{L}}{\partial \rho_d \partial \rho_w} \\ \bullet & \frac{\partial^2 \mathcal{L}}{\partial \rho_o^2} & \frac{\partial^2 \mathcal{L}}{\partial \rho_o \partial \rho_w} \\ \bullet & \bullet & \frac{\partial^2 \mathcal{L}}{\partial \rho_w^2} \end{bmatrix} = \begin{bmatrix} V_{dd} & V_{do} & V_{dw} \\ \bullet & V_{oo} & V_{ow} \\ \bullet & \bullet & V_{ww} \end{bmatrix},$$

where for  $j, k \in (d, o, w)$ ,  $V_{jk} = -\text{tr}(W_j A^{-1} W_k A^{-1}) - \frac{1}{\sigma^2} y' W_j' W_k y$ . The steps below derive the expression of  $V_{dd}$ , while all other elements can be found using the same arguments and only require to exchange the indexes.

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial \rho_d^2} &= \frac{\partial^2}{\partial \rho_d^2} (\ln |A|) + \frac{\partial^2}{\partial \rho_d^2} \left( -\frac{1}{2\sigma^2} (Ay - Z\delta)' (Ay - Z\delta) \right) \\ &= \frac{\partial}{\partial \rho_d} \left( \text{tr}(A^{-1} \frac{\partial}{\partial \rho_d} A) \right) + \frac{\partial^2}{\partial \rho_d^2} \left( -\frac{1}{2\sigma^2} y' A' Ay \right) \\ &= \frac{\partial}{\partial \rho_d} \left( -\text{tr}(A^{-1} W_d) \right) + \frac{\partial}{\partial \rho_d} \left( -\frac{1}{2\sigma^2} 2y' A' (-W_d) y \right) \\ &= -\text{tr}(-A^{-1} (-W_d) A^{-1} W_d) - \frac{y' W_d' W_d y}{\sigma^2} \\ &= -\text{tr}(W_d A^{-1} W_d A^{-1}) - \frac{y' W_d' W_d y}{\sigma^2} := V_{dd} \end{aligned}$$

### Block $H_{21}^{(b)}$

The Hessian block  $H_{21}^{(b)}$  given in (20) corresponds to the cross derivative of the likelihood function which we differentiate first with respect to  $\theta = (\delta' \sigma^2)'$  and then with respect to  $\rho$

$$H_{21}^{(b)} = \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \rho'} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \delta \partial \rho_d} & \frac{\partial^2 \mathcal{L}}{\partial \delta \partial \rho_o} & \frac{\partial^2 \mathcal{L}}{\partial \delta \partial \rho_w} \\ \frac{\partial^2 \mathcal{L}}{\partial \sigma^2 \partial \rho_d} & \frac{\partial^2 \mathcal{L}}{\partial \sigma^2 \partial \rho_o} & \frac{\partial^2 \mathcal{L}}{\partial \sigma^2 \partial \rho_w} \end{bmatrix} = \begin{bmatrix} -\frac{Z' L}{\sigma^2} \\ \delta' (Z' L) - \tau(\rho)' (y' L) \\ \sigma^4 \end{bmatrix},$$

where  $L = (y^{(2)} y^{(3)} y^{(4)})$  is based on the definition of  $y^{(t)}$ , for  $t = 2, 3, 4$  in (13). The equations below derive analytical expressions for the elements in the first column of above matrix.

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial \delta \partial \rho_d} &= \frac{\partial^2}{\partial \delta \partial \rho_d} \left( -\frac{1}{2\sigma^2} (Ay - Z\delta)' (Ay - Z\delta) \right) & \frac{\partial^2 \mathcal{L}}{\partial \sigma^2 \partial \rho_d} &= \frac{\partial^2}{\partial \sigma^2 \partial \rho_d} \left( -\frac{1}{2\sigma^2} (Ay - Z\delta)' (Ay - Z\delta) \right) \\ &= \frac{\partial^2}{\partial \delta \partial \rho_d} \left( -\frac{1}{2\sigma^2} (-2(Z\delta)' Ay) \right) & &= \frac{\partial^2}{\partial \sigma^2 \partial \rho_d} \left( -\frac{1}{2\sigma^2} (y' A' Ay - 2(Z\delta)' Ay) \right) \\ &= \frac{\partial}{\partial \delta} \left( \frac{(Z\delta)' (-W_d) y}{\sigma^2} \right) & &= \frac{\partial}{\partial \rho_d} \left( \frac{y' A' Ay - 2(Z\delta)' Ay}{2\sigma^4} \right) \\ &= \frac{-Z' W_d y}{\sigma^2} & &= \frac{-2(Ay)' W_d y + 2\delta' Z' W_d y}{2\sigma^4} \end{aligned}$$

The previously introduced permutation-invariance of the likelihood function then allows to apply the same steps to derive the other two columns. These expressions are simplified further using the multivariate formulation  $L = (y^{(2)} y^{(3)} y^{(4)})$  of the spatial lags of the flow vector.

$$\begin{aligned}
H_{21}^{(b)} &= \begin{bmatrix} \frac{-Z'W_d y}{\sigma^2} & \frac{-Z'W_o y}{\sigma^2} & \frac{-Z'W_w y}{\sigma^2} \\ \frac{\delta'Z'W_d y - (Ay)'W_d y}{\sigma^4} & \frac{\delta'Z'W_o y - (Ay)'W_o y}{\sigma^4} & \frac{\delta'Z'W_w y - (Ay)'W_w y}{\sigma^4} \end{bmatrix} \\
&= \begin{bmatrix} \frac{-Z'y^{(2)}}{\sigma^2} & \frac{-Z'y^{(3)}}{\sigma^2} & \frac{-Z'y^{(4)}}{\sigma^2} \\ \frac{\delta'Z'y^{(2)} - \tau(\rho)'y_\bullet y^{(2)}}{\sigma^4} & \frac{\delta'Z'y^{(3)} - \tau(\rho)'y_\bullet y^{(3)}}{\sigma^4} & \frac{\delta'Z'y^{(4)} - \tau(\rho)'y_\bullet y^{(4)}}{\sigma^4} \end{bmatrix} \\
&= \begin{bmatrix} \frac{-Z'L}{\sigma^2} \\ \frac{\delta'(Z'L) - \tau(\rho)'(y_\bullet L)}{\sigma^4} \end{bmatrix}
\end{aligned}$$

### Block $H_{22}^{(b)}$

The Hessian block  $H_{22}^{(b)}$  given in (20) contains the second order derivatives of the likelihood function with respect to  $\theta$

$$H_{22}^{(b)} = \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta'} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \delta \partial \delta'} & \frac{\partial^2 \mathcal{L}}{\partial \delta \partial \sigma^2} \\ \frac{\partial^2 \mathcal{L}}{\partial \sigma^2 \partial \delta'} & \frac{\partial^2 \mathcal{L}}{\partial (\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} -\frac{Z'Z}{\sigma^2} & 0 \\ 0 & -\frac{N}{2\sigma^4} \end{bmatrix}$$

When we evaluate the above block at the values of the maximum likelihood estimators in (16), in particular  $\hat{\sigma}^2(\rho) = (Ay - Z\delta(\rho))'(Ay - Z\delta(\rho))N^{-1}$  and  $\hat{\delta}(\rho) = (Z'Z)^{-1}Z' Ay$ , it has the same expression as the Hessian matrix of a non-spatial linear model.

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \delta \partial \delta'} &= \frac{\partial^2}{\partial \delta \partial \delta'} \left( -\frac{1}{2\sigma^2} (Ay - Z\delta)' (Ay - Z\delta) \right) \\
&= \frac{\partial^2}{\partial \delta \partial \delta'} \left( \frac{-\delta' Z' Z \delta}{2\sigma^2} \right) \\
&= \frac{-Z' Z}{\sigma^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \delta \partial \sigma^2} &= \frac{\partial^2}{\partial \delta \partial \sigma^2} \left( -\frac{1}{2\sigma^2} (Ay - Z\delta)' (Ay - Z\delta) \right) \\
&= \frac{\partial}{\partial \delta} \left( \frac{1}{2\sigma^4} (Ay - Z\delta)' (Ay - Z\delta) \right) \\
&= \frac{-Z' (Ay - Z\delta)}{\sigma^4} \\
&= \frac{-Z' Ay - Z' Z (Z' Z)^{-1} Z' Ay}{\sigma^4} \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial (\sigma^2)^2} &= \frac{\partial^2}{\partial (\sigma^2)^2} \left( -\frac{N}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (Ay - Z\delta)' (Ay - Z\delta) \right) \\
&= \frac{\partial}{\partial \sigma^2} \left( -\frac{N}{2\sigma^2} + \frac{(Ay - Z\delta)' (Ay - Z\delta)}{2\sigma^4} \right) \\
&= \frac{N}{2\sigma^4} - \frac{(Ay - Z\delta)' (Ay - Z\delta)}{\sigma^6} \\
&= \frac{N}{2\sigma^4} - \frac{N\sigma^2}{\sigma^6} \\
&= -\frac{N}{2\sigma^4}
\end{aligned}$$

## Appendix B A matrix formulation of spatial instruments

This appendix shows how to derive the moments  $U'y$ ,  $U'L$  and  $U'U$  of Section 4.4.2 that are required to perform S2SLS estimation based on the matrix formulation of the spatial interaction model. The first subsection presents the ten blocks of the variance moment  $U'U$  and the following one derives the covariance moments  $U'y$  and  $U'L$ .

### The variance moment $U'U$

The previously introduced groups of instrumental variables  $U = (U_\alpha U_{\alpha_I} U_\beta U_\gamma)$  allow to express the moment matrix  $U'U$  in terms of 16 blocks.

$$U'U = \begin{pmatrix} U_{\alpha\alpha} & U_{\alpha\alpha_I} & U_{\alpha\beta} & U_{\alpha\gamma} \\ \bullet & U_{\alpha_I\alpha_I} & U_{\alpha_I\beta} & U_{\alpha_I\gamma} \\ \bullet & \bullet & U_{\beta\beta} & U_{\beta\gamma} \\ \bullet & \bullet & \bullet & U_{\gamma\gamma} \end{pmatrix}$$

The following four paragraphs derive the diagonal blocks of above matrix. For the six off-diagonal blocks we will start with those in the first row and move from top to bottom and left to

right. Since  $U'U$  is a symmetric matrix, we can focus on the upper triangle of the above matrix and infer all blocks on the lower triangle by symmetry.

### Block 1 ( $\alpha^2$ )

The first block of the  $U'U$  moment contains only the inner product of intercept term  $U_\alpha = \iota_N$ .

$$U'_\alpha U_\alpha := U_{\alpha\alpha} = (\iota'_n \otimes \iota'_n)(\iota_n \otimes \iota_n) = (\iota'_n \iota_n)(\iota'_n \iota_n) = n \times n = N$$

### Block 2 ( $\alpha_I^2$ )

The second diagonal block of the covariance moment  $U'_{\alpha_I} U_{\alpha_I} := U_{\alpha_I \alpha_I}$  contains the inner product of the instruments derived from intra-regional constant. All columns in  $U_{\alpha_I}$  are derived by applying VEC-operator to the matrices  $A_i^H$ , for  $i = 1, \dots, 9$ , which are defined below.

$$U_{\alpha_I} = (\text{VEC}(A_1^U) \quad \dots \quad \text{VEC}(A_9^U))$$

$$\begin{aligned} A_1^U &= \mathbf{I}_n & A_4^U &= W^2 & A_7^U &= W^2 W^{2'} \\ A_2^U &= W & A_5^U &= W^{2'} & A_8^U &= W^2 W' \\ A_3^U &= W' & A_6^U &= W W' & A_9^U &= W W^{2'} \end{aligned} \quad (50)$$

This structure allows to compute each element of  $U_{\alpha_I \alpha_I}$  as  $U_{\alpha_I \alpha_I, ij} = \iota'_n (A_i^H \odot A_i^U) \iota_n$ , for  $i, j = 1, \dots, 9$ . Since the block is symmetric we only need to compute the elements on the upper triangle. We can further reduce the number of elements to compute if we exploit the following properties: For two square matrices  $A$  and  $B$  with compatible dimension, we have  $\iota'_n (A \odot B) \iota_n = \iota'_n (A' \odot B') \iota_n = \iota'_n (B \odot A) \iota_n$ . In addition, if either  $A$  or  $B$  is symmetric we have  $\iota'_n (A \odot B) \iota_n = \iota'_n (A' \odot B) \iota_n$ . Exploiting all symmetries of the problem we only need to compute 21 of the 81 elements that constitute the matrix  $U_{\alpha_I \alpha_I}$  for the case when  $W$  itself is not symmetric. When  $W$  is symmetric we have to define  $U_\alpha$  in terms of the five instruments  $\text{VEC}(W^m)$  for  $m = 0, 1, 2, 3, 4$  since four of the nine vectors in  $(\text{VEC}(A_1^U) \quad \dots \quad \text{VEC}(A_9^U))$  would be redundant.

### Block 3 ( $\beta^2$ )

The next diagonal block contains the inner product of the instruments derived from the site attributes  $U_\beta = (\tilde{X}_d \tilde{X}_o \tilde{X}_I)$ . We can exploit the structure of the three matrices  $\tilde{X}_r$  ( $r = d, o, I$ ) to derive this moment block from matrix products of much the much smaller matrices  $\tilde{D}\tilde{X}$ ,  $\tilde{O}\tilde{X}$  and  $\tilde{I}\tilde{X}$ .

$$U'_\beta U_\beta := U_{\beta\beta} = \begin{pmatrix} n\tilde{D}\tilde{X}'\tilde{D}\tilde{X} & (\iota'_n \tilde{D}\tilde{X})'(\iota'_n \tilde{O}\tilde{X}) & \tilde{D}\tilde{X}'\tilde{I}\tilde{X} \\ \bullet & n\tilde{O}\tilde{X}'\tilde{O}\tilde{X} & \tilde{O}\tilde{X}'\tilde{I}\tilde{X} \\ \bullet & \bullet & \tilde{I}\tilde{X}'\tilde{I}\tilde{X} \end{pmatrix}$$

### Block 4 ( $\gamma^2$ )

The last diagonal block contains inner product of the instruments derived from the pair attributes  $U_\gamma = (\text{VEC}(G) \text{VEC}(\check{G}) \text{VEC}(\check{\check{G}}))$ , where  $\check{G} = WGW'$  and  $\check{\check{G}} = W\check{G}W'$ . This structure allows to express the moment block in terms of the Hadamard product. When the matrix  $G$  represents the geographic distance, it is usually symmetric, which implies that  $\check{G}$  and  $\check{\check{G}}$  are also symmetric.

In this case, we can exploit that for two symmetric matrices  $A$  and  $B$  with compatible dimensions, we only require the upper diagonal of both to compute  $\iota'_n(A \odot B)\iota_n$ . This reduces the memory and computational requirements by about 50%, which may be worthwhile because this block is responsible for a major part of the computational burden of the S2SLS estimator.

$$U'_\gamma U_\gamma := U_{\gamma\gamma} = \begin{pmatrix} \iota'_n(G \odot G)\iota_n & \iota'_n(G \odot \check{G})\iota_n & \iota'_n(G \odot \check{\check{G}})\iota_n \\ \bullet & \iota'_n(\check{G} \odot \check{G})\iota_n & \iota'_n(\check{G} \odot \check{\check{G}})\iota_n \\ \bullet & \bullet & \iota'_n(\check{\check{G}} \odot \check{\check{G}})\iota_n \end{pmatrix}$$

### Block 5 ( $\alpha\alpha_I$ )

The fifth block of  $U'U$  is obtained as the inner product of the global constant  $U_\alpha$  and the instruments derived from the intra-regional constant  $U_{\alpha_I}$ . Since all elements of  $U_\alpha$  are one, computing this inner product corresponds to summing the elements in each of the nine vectors in  $U_{\alpha_I}$ . As  $W$  is assumed to be row-stochastic, we can directly conclude that the first five entries of  $U_{\alpha\alpha_I}$  are equal to  $n$ . The remaining four entries are computed from the vectors of column sums of  $W$  and  $W^2$ .

$$U'_\alpha U_{\alpha_I} := U_{\alpha\alpha_I} = (n \quad n \quad n \quad n \quad n \quad (\iota'_n W)(\iota'_n W)' \quad (\iota'_n W^2)(\iota'_n W^2)' \quad (\iota'_n W)(\iota'_n W^2)' \quad (\iota'_n W)(\iota'_n W^2)')$$

### Block 6 ( $\alpha\beta$ )

Block six is computed from the constant  $U_\alpha$  and the instruments  $U_\beta$  that are derived from the site attributes. Computing the elements of this block only requires the scalar  $n$  and the column sums of the matrices  $\tilde{D}\tilde{X}$ ,  $\tilde{O}\tilde{X}$  and  $\tilde{I}\tilde{X}$ .

$$U'_\alpha U_\beta := U_{\alpha\beta} = \begin{pmatrix} n \cdot \iota'_n \tilde{D}\tilde{X} & n \cdot \iota'_n \tilde{O}\tilde{X} & \iota'_n \tilde{I}\tilde{X} \end{pmatrix}$$

### Block 7 ( $\alpha\gamma$ )

This block contains the inner product of the constant  $U_\alpha$  with the instruments derived from the exogenous attributes of the origin-destination pairs  $U_\gamma$ .

$$U'_\alpha U_\gamma := U_{\alpha\gamma} = \begin{pmatrix} \iota'_n G\iota_n & \iota'_n \check{G}\iota_n & \iota'_n \check{\check{G}}\iota_n \end{pmatrix}$$

### Block 8 ( $\alpha_I\beta$ )

The eighth block of  $U'U$  is the empirical covariance between  $U_{\alpha_I}$  the instruments derived from the intra-regional constant and  $U_\beta$  the instruments derived from the site attributes. The structure of  $U_{\alpha_I}$  allows to express each of the nine rows of  $U'_{\alpha_I} U_\beta := U_{\alpha_I\beta}$  using  $A_i^U$ , for  $i = 1, \dots, 9$ , as defined in (50).

$$U_{\alpha_I\beta,i} = \begin{pmatrix} (\iota'_n A_i^H) \tilde{D}\tilde{X} & (\iota'_n A_i^H) \tilde{O}\tilde{X} & \text{diag}(A_i^H) \tilde{I}\tilde{X} \end{pmatrix}$$

### Block 9 ( $\alpha_I\gamma$ )

This block is calculated as the inner product of the instruments derived from the intra-regional  $U_{\alpha_I}$  constant and those derived from the exogenous origin-destination pair attributes  $U_\gamma$ . It is again possible to derive  $U'_{\alpha_I} U_\gamma := U_{\alpha_I\gamma}$  for all rows ( $i = 1, \dots, 9$ ) using the notations in (50).

$$U_{\alpha_I \gamma, i} = \left( \iota'_n(A_i^H \odot G) \iota_n \quad \iota'_n(A_i^H \odot \check{G}) \iota_n \quad \iota'_n(A_i^H \odot \check{\check{G}}) \iota_n \right)$$

We can reduce the computational burden of this block if we pay attention to symmetries that were already mentioned in the sections on the diagonal blocks  $U_{\alpha_I \alpha_I}$  and  $U_{\gamma \gamma}$ .

### Block 10 ( $\beta \gamma$ )

The last block of the  $U'U$  moment matrix is computed as the inner product of  $U_\beta$  the instruments derived from the site attributes with  $U_\gamma$  the instruments derived from the pair attributes.

$$U'_{\beta \gamma} = \begin{pmatrix} (\iota'_n G') D \tilde{X} & (\iota'_n G) O \tilde{X} & \text{diag}(G) I \tilde{X} \\ (\iota'_n \check{G}') D \tilde{X} & (\iota'_n \check{G}) O \tilde{X} & \text{diag}(\check{G}) I \tilde{X} \\ (\iota'_n \check{\check{G}}') D \tilde{X} & (\iota'_n \check{\check{G}}) O \tilde{X} & \text{diag}(\check{\check{G}}) I \tilde{X} \end{pmatrix}$$

### The covariance moments $U'y_\bullet$

To construct the empirical covariances  $U'y$  and  $U'L$  we use the notations  $Y^{(t)}$  and  $y^{(t)}$  that are defined in (13). The elements of the four moments  $U'y^{(t)}$ , for  $t = 1, 2, 3, 4$ , represent the columns of  $U'y = U'y^{(1)}$  and  $U'L = (U'y^{(2)} U'y^{(3)} U'y^{(4)})$ .

$$U'y^{(t)} = \left( (U'_\alpha y^{(t)})' \quad (U'_{\alpha_I} y^{(t)})' \quad (U'_\beta y^{(t)})' \quad (U'_\gamma y^{(t)})' \right)'$$

The elements of the above moments are derived below, where the elementwise notation for the nine entries of  $U'_{\alpha_I} y^{(t)}$  uses definition (50) of  $A_i^U$ , for  $i = 1, \dots, 9$ .

$$U'_\alpha y^{(t)} = (\iota'_n Y^{(t)} \iota_n)$$

$$(U'_{\alpha_I} y^{(t)})_i = (\iota'_n (A_i^H \odot Y^{(t)}) \iota_n), \text{ for } i = 1, \dots, 9$$

$$U'_\beta y^{(t)} = \begin{pmatrix} D \tilde{X}' (Y^{(t)} \iota_n) \\ O \tilde{X}' (Y^{(t)} \iota_n) \\ I \tilde{X}' \text{diag}(Y^{(t)}) \end{pmatrix}$$

$$U'_\gamma y^{(t)} = \begin{pmatrix} \iota'_n (G \odot Y^{(t)}) \iota_n \\ \iota'_n (\check{G} \odot Y^{(t)}) \iota_n \\ \iota'_n (\check{\check{G}} \odot Y^{(t)}) \iota_n \end{pmatrix}$$