WORKING PAPERS

N° 1312

February 2022

"A generalized framework for estimating spatial econometric interaction models"

Lukas Dargel and Christine Thomas-Agnan

Toulouse School of Economics

A generalized framework for estimating spatial econometric interaction models

Lukas Dargel [1,2] and Christine Thomas-Agnan [1]

[1]Toulouse School of Economics, University of Toulouse Capitole, France
[2]BVA, France

February 28, 2022

Abstract

In the framework of spatial econometric interaction models for origin-destination flows, we develop an estimation method for the case when the list of origins may be distinct from the list of destinations, and when the origin-destination matrix may be sparse. The proposed model resembles a weighted version of the one of LeSage and Pace (2008) and we are able to retain most of the efficiency gains associated with the matrix form estimation, which we illustrate for the maximum likelihood estimator. We also derive computationally feasible tests for the coherence of the estimation results and present an efficient approximation of the conditional expectation of the flows.

Keywords: spatial econometric interaction models, zero flow problem

JEL Codes: C01 C21

1 Introduction

Spatial interaction models describe interaction behavior that occurs between a set of origins and a set of destinations. Some typical applications for this model are international trade flows between countries, passenger flows between cities, or geomarketing flows of customers who reside in the districts of a city and who visit the stores of a brand. Traditionally, this type of problem has been formulated in terms of the gravity equation, which assumes the size of the flow to increase in proportion to the mass of the origins and destinations and to decrease in proportion to the distance. Some examples for mass variables are the size of the population, or the surface area of a store. The success of the gravity equation can be explained by its intuitive simplicity and its high ability to explain the observed data in empirical applications. Also the criticism that these models describe macro-level behavior without foundation in individual actions has been overcome (see for example Wilson 1967; Bergstrand 1985). One legitimate concern, however, arises from the fact that most gravity models rely on independently distributed data to provide efficient and unbiased parameter estimates and predictions. In the context of interaction behavior, this independence assumption is usually untenable and has been refuted by empirical evidence in very diverse applications, such as public transport Kerkman, Martens, and Meurs (2018), air-passengers transport Margaretic, Thomas-Agnan, and Doucet (2017), home-to-work commuting Dargel (2021), international trade Fischer and LeSage (2020), remittances Laurent, Margaretic, and Thomas-Agnan (2020), migration Chun and Griffith (2011), or social interactions Wang et al. (2018).

LeSage and Pace (2008) propose a spatial econometric interaction model that explicitly models spatial dependence in origin-destination (OD) flows. Their model is particularly appealing because it can be estimated using a matrix representation of the flows, which reduces the dimension of most objects we need to manipulate during the estimation from N to $\sqrt{N}$, where N represents the number of OD pairs. For spatial econometric models this reduction may be critical in making the difference whether the model is computationally feasible in large data sets or not. However, to attain this efficiency gains the model of LeSage and Pace (2008) relies on two conditions that reduce its applicability to real-world data. The first condition requires that the set of origins coincides with the set of destinations and excludes, for example, geomarketing applications. The second condition requires that we actually observe interaction behavior for all possible OD pairs, which is unlikely in practice, as data sets with high spatial resolutions are increasingly frequent. In this article we present a generalization to the spatial econometric interaction models that relaxes both of the previously mentioned constraints, while maintaining most of the efficiency gains linked to the matrix form estimation. To do so, we develop a framework that treats flows as interactions between the nodes of spatial networks and derive new properties of Kronecker product that allow to include weights in the previous version of the matrix based estimator.

The previously mentioned issue of missing entries in the OD matrix is linked to the well known zero flow problem, as unobserved flows may be considered as implicitly reflecting zero values. When the number of such zeros is large the data is clearly not normally distributed, and alternative models, such as Tobit or zero inflated Poisson regression, have been proposed to accommodate the excess zeros. Burger, Oort, and Linders (2009), reviews and compares such methods in the context of international trade flows, but without specific focus on spatial dependence. Krisztin and Fischer (2015) extend the zero inflated Poisson model of trade flows by spatial filtering (Griffith 2003), which addresses the problem of biases in the parameter estimates due to spatial correlation. However, the main drawback of spatial filtering is that spatial dependence is treated as a nuisance, which means that we loose the ability to quantify spillover effects (Pace, Lesage, and Zhu 2013). There exist models that simultaneously account for excess zeros and explicitly model spatial dependence, such as the Tobit models put forward by LeSage and Pace (2009) and Xu and Lee (2015) or the family of models for spatially correlated limited dependent variables developed in Liesenfeld, Richard, and Vogler (2016), but it is clear that computational constraints make these approaches less feasible for big data environments. Unlike the previous methods, the model presented here maintains the hypotheses of gaussian data by only considering the subsample of OD pairs related to observed flows. It is easy to appreciate the computational advantage of this procedure, but we have to be aware that the obtained results do not generalize to the unobserved OD pairs. This would only be possible if the flows were missing at random, which is hard to justify in our context as unobserved flows are most often linked with high distances. While, this is certainly a drawback of the model here, it might be the better compromise if the alternative is ignoring spatial dependence altogether. Moreover, Linders and Groot (2006) find that the omission of zero flows yields results surprisingly similar to those of a model that accounts for sample selection, and much better than Tobit models or gaussian models in which the unobserved values are imputed by a constant.

The structure of the article is as follows: the next section introduces the origin-destination flow problem from the perspective of interactions between networks and develops the generalized version of the model. Section 3 presents the matrix form estimation for the MLE. The final section concludes.

2 A generalized framework for spatial interaction models

In this section we present origin-destination flows from the perspective of pairwise interactions between the nodes of an origin network and those of a destination network. Our framework uses the generalizations of Laurent, Margaretic, and Thomas-Agnan (2020) to cover both: the case of interactions within the same network, henceforth the square case, and the case of interactions between two distinct networks, labeled the rectangular case. An example of the square case is international trade, where the origin network as well as the destination network correspond to the countries of the world. In addition to the distinction between rectangular and square, we differentiate the cartesian from the non-cartesian case. The former applies when all possible interactions are actually observed, implying that the number OD pairs correspond to the cartesian product of all origins with all destinations. When some of the possible OD pairs are unobserved we are in the non-cartesian case.

Notations and definition of cases

π_o: origin network

π_d: destination network

n_o: number of nodes in the origin network

n_d: number of nodes in the destination network

OW: $n_o \times n_o$ neighborhood matrix of the origin network

DW: $n_d \times n_d$ neighborhood matrix of the destination network

$\mathcal{F}$: set of all potential OD pairs $\mathcal{F} = \{(o_j \to d_i) : j = 1, 2, ..., n_o, i = 1, 2, ..., n_d\}$

Γ: the subset of OD pairs for which we observe interaction data $\Gamma \subseteq \mathcal{F}$

N: total number of OD pairs $N = n_o \cdot n_d = \text{card}(\mathcal{F})$, where card is the cardinality of a set

N^*: number of OD pairs included in the model $N^* = \text{card}(\Gamma)$

Using the above notations we formalize the distinction if our cases. The **square case** treats flows within the same network $\pi_o = \pi_d$, which implies $n_o = n_d$ and $OW = DW$. In the **rectangular case** we treat flows between two distinct networks $\pi_o \neq \pi_d$. The **cartesian case** applies when we model all theoretically possible OD pairs $\Gamma = \mathcal{F}$, yielding $N^* = n_o \cdot n_d = N$. In contrast, the **non-cartesian case**, applies when we model a strict subset of all potential OD pairs $\Gamma \subset \mathcal{F}$, which leads to $N^* < n_o \cdot n_d = N$.

An illustrative example

To illustrate the different cases let us consider two networks π_1 and π_2, whose number of nodes are n_1 and n_2. The matrix $\mathbf{Y}_\diamond$, with dimensions $(n_2 + n_1) \times (n_2 + n_1)$, represents all possible pairwise interactions between nodes that belong to any of the two networks. We may then group these interactions into four sub-matrices

$$\mathbf{Y}_\diamond = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix}, \tag{1}$$

where the flows within the network π_1 are represented by $Y_{11}(n_1 \times n_1)$ and those within the network π_2 by $Y_{22}(n_2 \times n_2)$. Similarly, flows that connect the nodes of π_2 with those of π_1 are contained in $Y_{12}(n_1 \times n_2)$, and the flows connecting the nodes both networks in the other direction in $Y_{21}(n_2 \times n_1)$. Hence, if we model a diagonal blocks in (1) we are in the square case and for the off-diagonal ones we are in the rectangular case. Whether we observe all values in the sub-matrix then defines if the case is cartesian or not .

Differentiating the square from the rectangular case is trivial when the separation of nodes into two distinct networks is already given, but in practice such considerations may be up to the definition of the researcher. Our advise regarding this issue is to consider the potential neighborhood links between all of the observations. Separating origins and destinations into two networks is appropriate if neighborhood relations within each subset are conceivable but not between the two. Hence, in our example, the neighborhood matrix of all observations should have the form

$$\mathbf{W}_\diamond = \begin{pmatrix} W_1 & \mathbf{0} \\ \mathbf{0} & W_2 \end{pmatrix}, \tag{2}$$

where the matrices $W_1(n_1 \times n_1)$ and $W_2(n_2 \times n_2)$ capture the neighborhood relations between the nodes the networks π_1 and π_2. When it is not possible to defend the zero matrices on the off-diagonal blocks in (2) we should probably treat all observations as part of the same network and consider the case as square. In geomarketing applications the separation is apparent from the fact that the origins (residential areas) are conceptually different from the destinations (stores). When this conceptual distinction is not possible we could also use geographical arguments for such a separation as for example when modeling investment flows from the United States to the Chinese provinces.

2.1 The cartesian model

To define the spatial econometric interaction model we need to fix the role of the origin and destination networks. Within the formalism of the previous example this choice corresponds to setting $\pi_o = \pi_j$ and $\pi_d = \pi_i$ for one pair of $i, j = 1, 2$. Given the roles of each network we may extract the part of the interaction matrix we want to model $Y = Y_{ij}$ and define the neighborhood matrix of the origins $OW = W_j$ and that of the destinations $DW = W_i$. These neighborhood matrices OW and DW should have only non-negative entries, a zero diagonal, be similar to a symmetric matrix and have spectral radius of one. In practice, these conditions are met if we use matrices based on pairwise distances or contiguity, and normalize them by row or with respect to their spectral radius. When all OD pairs are included in the model we may use the Kronecker product $\otimes$ to derive three OD neighborhood matrices from the node-level neighborhood matrices.

$$\underset{N\times N}{W_d} = I_{n_o} \otimes DW \qquad \underset{N\times N}{W_o} = OW \otimes I_{n_d} \qquad \underset{N\times N}{W_w} = OW \otimes DW \tag{3}$$

In the model W_o represents origin-based dependence, W_d represents destination-based dependence and W_w represents origin-to-destination dependence. Definition (3) of these three matrices makes clear that if we model interaction within the same network ($\pi_o = \pi_d$), we find that $OW = DW := W$, which recovers the traditional framework of LeSage and Pace (2008). We use these three matrices in conjunction with the three autoregression parameters $\rho = \begin{pmatrix}\rho_d & \rho_o & \rho_w\end{pmatrix}$ to define the spatial filter matrix ($A = I_N - \rho_d W_d - \rho_o W_o - \rho_w W_w$). This matrix can be used to remove spatial autocorrelation from the flow vector $y = \text{Vec}(Y)$, which we obtain by stacking the columns of the flow matrix. If we relate the spatially filtered flows to linear combination of some exogenous variables $Z(N \times K)$ and a gaussian error $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_N)$

$$Ay = Z\delta + \varepsilon \tag{4}$$

we obtain a spatial lag (LAG) model. In Section 3 we present the Z matrix in details, but for the introduction of the model in its vectorized form the current definition suffices. It is also important to note that the above model is only well defined if the inverse of the filter matrix exists, an issue that will be treated in Section 3.3.

2.2 The non-cartesian model

Model (4) has the disadvantage that we can only use it when we actually observe all values in the flow matrix. In practice, this condition is rarely fulfilled, especially if we want to model flow data with a high spatial resolution. To clarify this issue let us consider an origin network with three nodes $\pi_o = \{o_1, o_2, o_3\}$ and a destination network with two nodes $\pi_d = \{d_1, d_2\}$, where flows from o_1 to d_2 and from o_3 to d_1 are theoretically possible but unobserved.

$$Y = \begin{pmatrix} y_{o_1\to d_1} & y_{o_2\to d_1} & \times \\ \times & y_{o_2\to d_2} & y_{o_2\to d_2} \end{pmatrix}, \tag{5}$$

It may be tempting to replace missing entries in the flow matrix by zeros and go on with model (4), but this would introduce a point-mass at zero, invalidating the normality assumption. To avoid this inconsistency, we prefer to drop the unobserved flows and create a truncated flow vector of the form

$$y^* = \begin{pmatrix} y_{o_1\to d_1} & y_{o_2\to d_1} & y_{o_2\to d_2} & y_{o_3\to d_2} \end{pmatrix}'. \tag{6}$$

This truncated flow vector only contains the subset of OD pairs we want to include in the model. We then define the part of the spatial filter matrix $A^*(N^* \times N^*)$ and the explanatory variables $Z^*(N^* \times K)$ that correspond to the same subset of OD pairs to define a model for the truncated sample

$$A^* y^* = Z^* \delta + \varepsilon^*, \tag{7}$$

where the error is supposed to be gaussian $\varepsilon^* \sim \mathcal{N}(0, \sigma^2 I_{N^*})$. We can link the matrices of the above, non-cartesian model, to those of the cartesian model in (4), by means of the selection operator $S_\Gamma(N \times N^*)$. The effect of this operator on a matrix or vector is to drop the columns or rows at the position of the OD pairs, for which we do not observe interaction data.

$$A^* = S'_\Gamma A S_\Gamma \qquad Z^* = S'_\Gamma Z \qquad y^* = S'_\Gamma y \tag{8}$$

For the example considered in (5) and (6) the selection operator and the flow vector $y = \text{Vec}(Y)$ would be

$$S_\Gamma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad y = \begin{pmatrix} y_{o_1 \to d_1} \\ \times \\ y_{o_2 \to d_1} \\ y_{o_2 \to d_2} \\ \times \\ y_{o_2 \to d_2} \end{pmatrix}.$$

It is easy to verify that the selection operator satisfies $S'_\Gamma S_\Gamma = I^*_N$ and $S_\Gamma S'_\Gamma = R_\Gamma$, where R_Γ replaces the unobserved flows in y with zeros instead of removing them. We can derive this replacement operator explicitly from a binary matrix of the observed flows, which is equal to one if a flow is observed and zero if it is not. Denoting this flow indicator matrix $\mathcal{I}_\Gamma$ the relationship is $R_\Gamma = \text{Diag}(\text{Vec}(\mathcal{I}_\Gamma))$, where the Diag operator places a vector on the main diagonal of a zero matrix. Below we illustrate these matrices for our previous example.

$$\mathcal{I}_\Gamma = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \qquad \text{Vec}(\mathcal{I}_\Gamma) = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \qquad R_\Gamma = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

3 Matrix form estimation in the general case

This section derives the matrix form estimator for the non-Cartesian model in (7). We focus on the maximum likelihood estimator (MLE), noting that extensions to the Bayesian Markov-Chain Monte-Carlo or spatial two-stage least squares estimators could be derived without much difficulty, using the arguments of Dargel (2021). The likelihood of the non-Cartesian model is

$$\mathcal{L}(\rho, \delta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{N^*/2} \exp\left\{\frac{1}{2\sigma^2}(A^*y^* - Z^*\delta)'(A^*y^* - Z^*\delta)\right\} |A^*|. \tag{9}$$

In the following four subsections we treat different parts of the estimation problem. The first one deals with methods that allow to efficiently evaluate the quadratic term $RSS(\rho, \delta) = (A^*y^* - Z^*\delta)'(A^*y^* - Z^*\delta)$. The next section deals with the log-determinant term $\log|A^*|$ that appears in the log-likelihood function. In the third subsection we discuss the issue of the feasible parameter space and in the last one we present a computationally efficient way to approximate the conditional expectation of the flows.

3.1 Moment calculation in matrix form

One key to the efficient estimation of the spatial econometric interaction model is to express the quadratic term in the likelihood function in terms of low dimensional moment matrices that are independent of the parameters. In what follows we first present some well-known and some new properties of Kronecker products, that allow to avoid computations with high-dimensional objects. Afterwards, we will treat the three moments that appear after expanding the quadratic term

$$RSS(\rho, \delta) = y^{*\prime} A^{*\prime} A^* y^* + \delta' Z^{*\prime} Z^* \delta - 2\delta' Z^{*\prime} A^* y^*. \tag{10}$$

3.1.1 Kronecker products, the Vec operator and the Diag operator

The following four properties are well known (Harville 1998) and have already been used to increase the efficiency of estimators for the cartesian spatial interaction model (see for example LeSage and Pace 2008; LeSage and Pace 2009): For three matrices A, B and C, whose dimensions allow to compute the matrix product ABC, and for two matrices D and E with identical dimensions $(n_a \times n_b)$ we have the following relations:

K1 : $(C' \otimes A)\,\text{Vec}(B) = \text{Vec}(ABC)$.
K2 : $\text{Diag}(\text{Vec}(D))\,\text{Vec}(E) = \text{Vec}(D) \odot \text{Vec}(E) = \text{Vec}(D \odot E)$
K3 : $\text{Vec}(D) + \text{Vec}(E) = \text{Vec}(D + E)$
K4 : $\text{Vec}(D)'\,\text{Vec}(E) = \iota_{n_a}'(D \odot E)\iota_{n_b}$

To improve the estimation efficiency in the non-Cartesian case we use three additional properties, which we formally derive in Appendix A. **Proposition:** Given four matrices $A(n_a \times k_a), B(n_b \times k_a), C(n_a \times n_b), D(n_a \times n_b)$ we have the following relations:

K5 : $(\iota_{n_b}' \otimes A')\,\text{Diag}(\text{Vec}(D))(B \otimes \iota_{n_a}) = A'DB$
K6 : $(\iota_{n_b}' \otimes A')\,\text{Diag}(\text{Vec}(D))(\iota_{n_b} \otimes A) = A'\,\text{Diag}(\iota_{n_b} D')A$
K7 : $(B' \otimes \iota_{n_a}')\,\text{Diag}(\text{Vec}(D))(B \otimes \iota_{n_a}) = B'\,\text{Diag}(\iota_{n_a} D)B$

3.1.2 The TSS moment

We refer to the first part of the RSS term in (10) as TSS-moment because it offers an analogy to the analysis of variance formula. Using $\tau(\rho)' = \begin{pmatrix} 1 & -\rho_d & -\rho_o & -\rho_w \end{pmatrix}$, and $y_\bullet = \begin{pmatrix} R_\Gamma y & W_d R_\Gamma y & W_o R_\Gamma y & W_w R_\Gamma y \end{pmatrix}$, we can factorize its expression into

$$y^{*\prime} A^{*\prime} A^* y^* = \tau(\rho)' y_\bullet' R_\Gamma y_\bullet \tau(\rho). \tag{11}$$

The 4×4 matrix in the middle of the factorization $TSS_\bullet = y_\bullet' R_\Gamma y_\bullet$ only depends on the data. Hence, it needs to be computed once and can then be used to quickly reevaluate the term for changing values of ρ. Based on the definition of the OD-level neighborhood matrices in (3) and the properties K1, K2, K4 we compute the elements of the $TSS_\bullet$ matrix for $l, k = 1, 2, 3, 4$ as

$$TSS_{\bullet kl} = \iota_{n_d}'(Y^{(k)} \odot \mathcal{I}_\Gamma \odot Y^{(l)})\iota_{n_o}, \tag{12}$$

where $Y^{(1)} = (\mathcal{I}_\Gamma \odot Y), Y^{(2)} = DW(\mathcal{I}_\Gamma \odot Y), Y^{(3)} = (\mathcal{I}_\Gamma \odot Y)OW'$ and $Y^{(4)} = DW(\mathcal{I}_\Gamma \odot Y)OW'$.

3.1.3 The variance moment

The variance moment $Z^{*\prime} Z^* = Z' R_\Gamma Z$ is proportional to the empirical variance of the explanatory variables. To exploit the benefits of the matrix form estimation we have to pay attention to the structure of the variables that are contained in the matrix Z. For our model we suppose that $Z = \begin{pmatrix} \iota_N & X_d & X_o & g \end{pmatrix}$ is composed of four sets of variables [1], where ι_N is a constant, X_d contains the characteristics of the origins, X_o those of the destinations and g is a vector characteristics for OD-pairs. In classical gravity models, this g vector reflects the geographic distance, but we could use much broader and also multiple measures of separation between origins and destinations. The variable in Z variables can be expressed as functions of the network-level data $OX(n_o \times k_o)$, $DX(n_d \times k_d)$ and a matrix representation of the OD-characteristics $G(n_d \times n_o)$.

$$\begin{aligned} \iota_N &= \iota_{n_o} \otimes \iota_{n_d} & X_d &= \iota_{n_o} \otimes DX \\ g &= \mathrm{Vec}(G) & X_o &= OX \otimes \iota_{n_d} \end{aligned} \tag{13}$$

Given the above structure of the matrices in Z and the definition of the replacement operator $R_\Gamma = \mathrm{Diag}(\mathrm{Vec}(\mathcal{I}_\Gamma))$, we can use the Kronecker product properties of Section 3.1.1 to derive the elements of the variance moment as

$$Z^{*\prime} Z^* = \begin{bmatrix} N^* & \iota'_{n_o} \mathcal{I}'_\Gamma DX & \iota'_{n_d} \mathcal{I}_\Gamma OX & \iota'_{n_d} (\mathcal{I}_\Gamma \odot G) \iota_{n_o} \\ \bullet & DX' \,\mathrm{Diag}(\iota'_{n_o} \mathcal{I}'_\Gamma) DX & DX' \mathcal{I}_\Gamma OX & DX' (\mathcal{I}_\Gamma \odot G) \iota_{n_o} \\ \bullet & \bullet & OX' \,\mathrm{Diag}(\iota_{n_d} \mathcal{I}_\Gamma) OX & OX' (\mathcal{I}'_\Gamma \odot G') \iota_{n_d} \\ \bullet & \bullet & \bullet & \iota'_{n_d} (G \odot \mathcal{I}_\Gamma \odot G) \iota_{n_o} \end{bmatrix}. \tag{14}$$

3.1.4 The covariance moments

The covariance moment $Z' R_\Gamma y_\bullet$ is proportional to the empirical covariances of the explanatory variables and the spatial lags of the flow vector. This moment appears when we use the notations in (11) to factor out the autocorrelation parameters from the third part of the RSS term in (10).

$$2\delta' Z^{*\prime} A^* y^* = 2\delta' Z' R_\Gamma y_\bullet \tau(\rho)$$

In the following we reuse definition (12) of $Y^{(t)}$, for $t = 1, 2, 3, 4$ to derive $y^{(t)} = \mathrm{Vec}(Y^{(t)})$, and to compute the elements of each column of the moment $Z' R_\Gamma y_\bullet$ as

$$Z' R_\Gamma y^{(t)} = \begin{bmatrix} \iota_{n_d} (\mathcal{I}_\Gamma \odot Y^{(t)}) \iota_{n_o} \\ DX' (\mathcal{I}_\Gamma \odot Y^{(t)}) \iota_{n_o} \\ OX' (\mathcal{I}'_\Gamma \odot Y^{*(t)\prime}) \iota_{n_d} \\ \iota'_{n_d} (G \odot \mathcal{I}_\Gamma \odot Y^{(t)}) \iota_{n_o} \end{bmatrix}. \tag{15}$$

3.2 Determinant calculation

In this section we treat the problem of evaluating the determinant term that appears in the likelihood function in (9). Evaluating this term based on classical decomposition methods such as the LU, QR, or Cholesky factorization would require $\mathcal{O}(N^{*3})$ operations, which may be prohibitive in large sample applications. This issue is well known in the spatial econometrics literature, and to address it we adapt an existing approximation method to model (7). The underlying idea was first proposed by Martin (1992) and later adjusted to the spatial econometric interaction model by LeSage and Pace (2008) and Fischer and LeSage (2020), who use it in the quadratic and Cartesian cases. We will first generalize this method to rectangular flows and then treat the non-cartesian case.

[1] Many extensions to the spatial econometric interaction model correspond to additional sets of variables in the matrix Z. We may, for example, use the spatial lags $W_d X_d$ and $W_o X_o$ as additional variables, to extend the LAG model considered in this article to a Spatial Durbin model.

The general form of the approximation is based on a Taylor series expression of the term $\ln|A^*| = \ln|I_{N^*} - W_F^*|$, where for the model considered in this article $W_F^* = S'_\Gamma W_F S_\Gamma$ and $W_F = \rho_d W_d + \rho_o W_o + \rho_w W_w$.

$$\ln|A^*| = -\sum_{t=1}^{\infty} \frac{\text{tr}\,(W_F^{*t})}{t}, \tag{16}$$

If all eigenvalues of W_F^* are less then one in magnitude the series in (16) converges and we can use the first m terms to approximate the log-determinant. One big advantage of this approximation results from the linearity of the trace operator, as we can factor out the parameters of the large matrix products required to compute $\text{tr}(W_F^{*t})$. This means that repeated evaluation of the log-determinant do not impact the computation time much when the traces have been already computed. To speed up the computation of the trace values we may further use the fact that the three matrices W_d, W_o and W_w represent a commuting family. In other words, the order of the matrices in the product is irrelevant and we can use a multinomial expansion to compute the values of the traces of W_F^t for $t = 1, 2, .., m$ by

$$\text{tr}(W_F^t) = \sum_{k_d+k_o+k_w=t} \binom{t}{k_d + k_o + k_w} \left(\rho_d^{k_d} \rho_o^{k_o} \rho_w^{k_w}\right) tr\left(W_d^{k_d} W_o^{k_o} W_w^{k_w}\right). \tag{17}$$

By exploiting the fact that $W_o W_d = W_d W_o = W_w$ we can further simplify the expression of the trace of the matrix product

$$tr(W_d^{k_d} W_o^{k_o} W_w^{k_w}) = tr(W_d^{k_d+k_w} W_o^{k_o+k_w}). \tag{18}$$

The Kronecker product structures of $W_d = (I_{n_o} \otimes DW)$ and $W_o = (OW \otimes I_{n_d})$ then allows to express the above trace in terms of the the traces of powers of the site neighborhood matrices.

$$\text{tr}(W_d^{k_d+k_w} W_o^{k_o+k_w}) = \text{tr}(OW^{k_o+k_w})\,\text{tr}(DW^{k_d+k_w}) \tag{19}$$

Unfortunately, the matrices W_d^*, W_o^*, and W_w^* lose their Kronecker products structure and generally do not commute $W_d^* W_o^* \neq W_d^* W_o^*$ in the non-Cartesian case. Consequently, we cannot apply the developments presented in the last three formulas and we are obliged to develop the terms $\text{tr}(W_F^{*t}) = \text{tr}((\rho_d W_d^* + \rho_o W_o^* + \rho_w W_w^*)^t)$ for $t = 1, 2, ..., m$, explicitly. The direct expansion would lead to $1^3 + 2^3 + ... + m^3$ products of $N^* \times N^*$ matrices, which is problematic as these matrices become increasingly dense. In Appendix B we develop some simplifications for the expansion required for the forth order approximation. These reduce the number matrix products to compute from 120 to 10 and remains more feasible in high dimensional applications.

3.3 Considerations about the feasible parameter space

The issue of the feasible parameter space is recurrent in the spatial econometrics literature and relates to the fact that model (7) is incoherent when the spatial filter matrix A^* is singular. This condition imposes constraints on the autoregression parameters. Dargel (2021) discusses this issue in the context of the cartesian and square flows and develops an efficient method to check the coherence of the model. In the following paragraphs we extend this method first to the rectangular and then to the non-cartesian model.

For the cartesian and rectangular cases, the filter matrix is given by $A = I_N - W_F$, with $W_F = \rho_d(I_{n_o} \otimes DW) + \rho_o(OX \otimes I_{n_d}) + \rho_w(OW \otimes DW)$. The necessary and sufficient condition for A to be non singular is that all its eigenvalues are different from zero. As this condition is to hard to work with in practice, we use the sufficient, but more restrictive alternative that the spectral radius $r(W_F)$ is smaller than one. Using the developments in Appendix C we can write the eigenvalues vector of $\lambda(W_F)$ of W_F as

$$\lambda(W_F) = \rho_d(\iota_{n_o} \otimes \lambda(DW)) + \rho_o(\lambda(OW) \otimes \iota_{n_d}) + \rho_w(\lambda(OW) \otimes \lambda(DW)). \tag{20}$$

This representation clearly shows that we can test the condition $r(W_F) < 1$ without having to construct W_F explicitly. Furthermore, to infer $r(W_F)$, we do not require to compute the full spectrum of OW and DW, but only their smallest and largest eigenvalue. We refer to these two extreme eigenvalues as λ_{max} and λ_{min}. When OW and DW are sparse these eigenvalues can be computed with only $\mathcal{O}(n_d) + \mathcal{O}(n_o)$ operations, using for example the implicitly restarted Arnoldi method of Sorensen (1992) or the Krylow-Schur algorithm of Stewart (2002).

It is obvious that equation (20) does not hold in the non-cartesian case, as we lose the Kronecker product structure of the spatial filter matrix A^*. Consequently, the eigenvalues of A^* cannot be directly expressed as a function of those of the matrices OW and DW. However, we can show that the previous test for non-singularity of A still allows to conclude that the non-cartesian model is coherent. To do so, let us first recall the definition of the spatial filter matrix in the non-cartesian case $A^* = I_{N^*} - W_F^*$, where $W_F^* = S_\Gamma' W_F S_\Gamma$ is clearly a principal sub-matrix of W_F, which allows to conclude that

$$\lambda_{min}(W_F) \leq \lambda_{min}(W_F^*) \qquad \text{and} \qquad \lambda_{max}(W_F^*) \leq \lambda_{max}(W_F). \tag{21}$$

When W_F is symmetric we may recursively apply Cauchy's interlacing theorem (Horn and Johnson 2012, page 242) to derive the above results. In Appendix C.2, we demonstrate that the same can also be shown for the more general form of the W_F matrix that was introduced in Section 2.1. The inequalities (21) clearly show that constraints placed on the spectral radius of W_F are more binding than those placed on the spectral radius of W_F^*. Hence, a condition that ensures the coherence of the cartesian model remains sufficient for the coherence of non-cartesian model.

3.4 Approximating the conditional expectation

Another practical concern is the computation of the conditional expectation of the value of the flows $\mathbb{E}[y^*|Z^*] = A^{*-1}Z^*\delta$. We require this expectation to compute predictions (see for example Goulard, Laurent, and Thomas-Agnan 2017) and to evaluate the impact measures of LeSage and Thomas-Agnan (2015). Since the direct computation of A^{*-1} is often not feasible we will use an approximation based on a series expression of the inverse matrix. For a standard spatial model this approximation is already suggested by LeSage and Pace (2009, page 111) and we will adapt it to the spatial interaction model and derive is matrix form expression.

Let us first recall that the inverse of a matrix may be derived as $A^{*-1} = \sum_{t=0}^{\infty}(I - A^*)^t$, which is a converging series if the spectral radius of A^* is inferior to one. This expression allows to approximate the conditional expectation without the need to compute an inverse.

$$\mathbb{E}[y^*|Z^*] \approx \sum_{t=1}^{m} (W_F^*)^t Z^* \delta$$

With some further restructuring we can also bypass computing the powers of W_F^*. This is done based on the recursive expression $z_o = Z\delta$ and $z_{t+1} = W_F R_\Gamma' z_t$, for $t = 0, 1, ..., m-1$, which we find by developing

$$\mathbb{E}[y^*|Z^*] \approx \sum_{t=0}^{m} (S_\Gamma' W_F S_\Gamma)^t S_\Gamma' Z\delta = S_\Gamma' \sum_{t=0}^{m} z_t,$$

We then denote by $\mathcal{Z}_t$ the matrix version of $z_t = \mathrm{Vec}(\mathcal{Z}_t)$, which allows further dimension reduction of our calculations.

$$\begin{aligned}
\mathrm{Vec}(\mathcal{Z}_{t+1}) &= W_F R_\Gamma \,\mathrm{Vec}(\mathcal{Z}_t) \\
&= (\rho_d W_d + \rho_o W_o + \rho_w W_w)\,\mathrm{Diag}(\mathrm{Vec}(\mathcal{I}_\Gamma))\,\mathrm{Vec}(\mathcal{Z}_t) \\
&= \mathrm{Vec}(\rho_d DW(\mathcal{I}_\Gamma \odot \mathcal{Z}_t) + \rho_o(\mathcal{I}_\Gamma \odot \mathcal{Z}_t)OW' + \rho_d DW(\mathcal{I}_\Gamma \odot \mathcal{Z}_t)OW')
\end{aligned}$$

The final approximation is then $\mathbb{E}[y^*|Z^*] \approx \sum_{t=0}^{m} S_\Gamma' \,\mathrm{Vec}(\mathcal{Z}_t)$ and for a fixed order m its quality depends on the values of the autoregressive parameters. When their magnitude is small the terms of the series will quickly tend to zero and for stronger autocorrelation the higher order terms become more important.

4 Conclusion

This article develops a new framework for estimating interaction models of spatially correlated origin-destination flows. We extend the approach of LeSage and Pace (2008) to allow for missing values in the OD matrix and to account for situations where the list of origins may different from the list of destinations. Our methodology allows to estimate the generalized model efficiently from a matrix representation of the flows, which we demonstrate for the MLE. With these generalizations, it is much easier to estimate the spatial econometric interaction model for increasingly common, high spatial resolution flow data. A limitation of the model is that the absence of a flow is not modeled explicitly, which means that, in general, the results apply only to the sub-sample of OD pairs with observed values for the flow. We leave extensions that simultaneously account for the selection mechanism and spatial autocorrelation for future research.

References

Bergstrand, Jeffrey H. (1985). “The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence”. In: *The Review of Economics and Statistics* 3, pp. 474–481.

Burger, Martijn, Frank van Oort, and Gert-Jan Linders (2009). “On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation”. In: *Spatial Economic Analysis* 2, pp. 167–190.

Chun, Yongwan and Daniel A. Griffith (2011). “Modeling Network Autocorrelation in Space–Time Migration Flow Data: An Eigenvector Spatial Filtering Approach”. In: *Annals of the Association of American Geographers* 3, pp. 523–536.

Dargel, Lukas (2021). “Revisiting estimation methods for spatial econometric interaction models”. In: *Journal of Spatial Econometrics* 1, p. 10.

Fischer, Manfred M. and James Paul LeSage (2020). “Network dependence in multi-indexed data on international trade flows”. In: *Journal of Spatial Econometrics* 1, 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* 2-3, pp. 304–325.

Griffith, Daniel A. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*. Springer Science & Business Media.
Harville, David A. (1998). “Matrix Algebra From a Statistician’s Perspective”. In: *Technometrics* 2, pp. 164–164.
Horn, Roger A. and Charles R. Johnson (2012). *Matrix Analysis*. Cambridge University Press.
Kerkman, Kasper, Karel Martens, and Henk Meurs (2018). “Predicting travel flows with spatially explicit aggregate models: On the benefits of including spatial dependence in travel demand modeling”. In: *Transportation Research Part A: Policy and Practice*, pp. 68–88.
Krisztin, Tamás and Manfred M. Fischer (2015). “The Gravity Model for International Trade: Specification and Estimation Issues”. In: *Spatial Economic Analysis* 4, pp. 451–470.
Laurent, Thibault, Paula Margaretic, and Christine Thomas-Agnan (2020). “Bilateral remittances and network effects: Do neighboring countries matter?” In:
LeSage, James Paul and Robert Kelley Pace (2008). “Spatial Econometric Modeling of Origin-Destination Flows”. In: *Journal of Regional Science* 5, pp. 941–967.
— (2009). *Introduction to Spatial Econometrics*. CRC Press.
LeSage, James Paul and Christine Thomas-Agnan (2015). “Interpreting Spatial Econometric Origin-Destination Flow Models”. In: *Journal of Regional Science* 2, pp. 188–208.
Liesenfeld, Roman, Jean-François Richard, and Jan Vogler (2016). “Likelihood Evaluation of High-Dimensional Spatial Latent Gaussian Models with Non-Gaussian Response Variables”. In: Emerald Group Publishing Limited, pp. 35–77.
Linders, Gert-Jan and Henri L. F. de Groot (2006). “Estimation of the Gravity Equation in the Presence of Zero Flows”. Tech. rep. ID 924160.
Margaretic, Paula, Christine Thomas-Agnan, and Romain Doucet (2017). “Spatial dependence in (origin-destination) air passenger flows”. In: *Papers in Regional Science* 2, pp. 357–380.
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* 1, pp. 189–205.
Pace, R. Kelley, James P. Lesage, and Shuang Zhu (2013). “Interpretation and Computation of Estimates from Regression Models using Spatial Filtering”. In: *Spatial Economic Analysis* 3, pp. 352–369.
Sorensen, D. C. (1992). “Implicit Application of Polynomial Filters in a k-Step Arnoldi Method”. In: *SIAM Journal on Matrix Analysis and Applications* 1, pp. 357–385.
Stewart, G. W. (2002). “A Krylov–Schur Algorithm for Large Eigenproblems”. In: *SIAM Journal on Matrix Analysis and Applications* 3, pp. 601–614.
Wang, Zheye et al. (2018). “A spatial econometric modeling of online social interactions using microblogs”. In: *Computers, Environment and Urban Systems*, pp. 53–58.
Wilson, A. G. (1967). “A statistical theory of spatial distribution models”. In: *Transportation Research* 3, pp. 253–269.
Xu, Xingbai and Lung-fei Lee (2015). “Maximum likelihood estimation of a spatial autoregressive Tobit model”. In: *Journal of Econometrics* 1, pp. 264–280.

Appendix A Additional properties of Kronecker products

In the following we provide the proofs for properties K5, K6 and K7, presented in Section 3.1.1. Before stepping into to these proofs it is useful to understand our indexing notation. Here we use $A_{[ij]}$ to access one element, $A_{[i\cdot]}$ access one row and $A_{[\cdot j]}$ to access one column of the matrix A. All three properties are defined for four matrices $A(n_a \times k_a), B(n_b \times k_a), C(n_a \times n_b), D(n_a \times n_b)$.

A.1 Proof of K5

To show the property $(\iota'_{n_b} \otimes A')\,\mathrm{Diag}(\mathrm{Vec}(D))(B \otimes \iota_{n_a}) = A'DB$ we exploit the block structure of the two Kronecker products $(\iota'_{n_b} \otimes A') = (A'\ A'\ \dots\ A')$ and $(B \otimes \iota_{n_a}) = ((B_{[1\cdot]} \otimes \iota_{n_a})\ (B_{[2\cdot]} \otimes \iota_{n_a})\ \dots (B_{[n_b\cdot]} \otimes \iota_{n_a}))$. Since the matrix $\mathrm{Diag}(\mathrm{Vec}(D))$ is diagonal we may break it down into blocks of size $(n_a \times n_a)$, where the diagonal blocks have the form $\mathrm{Diag}(D_{[\cdot k]})$, for $k = 1, ..., n_b$. This allows rewrite the overall product in terms of the sum of matrix multiplications for the individual blocks.

$$
\begin{aligned}
(\iota'_{n_b} \otimes A')\,\mathrm{Diag}(\mathrm{Vec}(D))(B \otimes \iota_{n_a}) &= \sum_{k=1}^{n_b} A'\,\mathrm{Diag}(D_{[\cdot k]})(B_{[k\cdot]} \otimes \iota_{n_a}) \\
&= A' \sum_{k=1}^{n_b} (B_{[k\cdot]} \otimes \mathrm{Diag}(D_{[\cdot k]})\iota_{n_a}) \\
&= A' \sum_{k=1}^{n_b} D_{[\cdot k]} B_{[k\cdot]}
\end{aligned}
$$

We conclude the proof by showing that the elements of the sum in the above expression are equal to those of the matrix product DB.

$$
(\sum_{k=1}^{n_b} D_{[\cdot k]} B_{[k\cdot]})_{[ij]} = \sum_{k=1}^{n_b} D_{[ik]} B_{[kj]} = (DB)_{[ij]} \quad \square
$$

A.2 Proof of K6

This property directly follows from the following development.

$$
\begin{aligned}
(\iota'_{n_b} \otimes A')\,\mathrm{Diag}(\mathrm{Vec}(D))(\iota_{n_b} \otimes A) &= A' \sum_{k=1}^{n_b} \mathrm{Diag}(D_{[\cdot k]})A \\
&= A'\,\mathrm{Diag}(\iota_{n_b} D')A \quad \square
\end{aligned}
$$

A.3 Proof of K7

The property $(B' \otimes \iota'_{n_a})\,\mathrm{Diag}(\mathrm{Vec}(D))(B \otimes \iota_{n_a}) = B'\,\mathrm{Diag}(\iota_{n_a} D)B$ is also derived from a a block-wise decomposition of the overall matrix product, but unlike before it is not possible to factor out these blocks.

$$
\begin{aligned}
(B' \otimes \iota'_{n_a})\,\mathrm{Diag}(\mathrm{Vec}(D))(B \otimes \iota_{n_a}) &= \sum_{k=1}^{n_b} (B'_{[k\cdot]} \otimes \iota'_{n_a})\,\mathrm{Diag}(D_{[\cdot k]})(B_{[k\cdot]} \otimes \iota_{n_a}) \\
&= \sum_{k=1}^{n_b} (B'_{[k\cdot]} B_{[k\cdot]} \otimes \iota'_{n_a}\,\mathrm{Diag}(D_{[\cdot k]})\iota_{n_a}) \\
&= \sum_{k=1}^{n_b} B'_{[k\cdot]} B_{[k\cdot]} \iota'_{n_a} D_{[\cdot k]}
\end{aligned}
$$

We conclude the proof by showing that demonstrating the equality of all elements in the above sum with those of $B' \operatorname{Diag}(\iota_{n_a} D)B$.

$$\left(\sum_{k=1}^{n_b} B'_{[k\cdot]} B_{[k\cdot]} \iota'_{n_a} D_{[\cdot k]}\right)_{[ij]} = \sum_{k=1}^{n_b} B_{[ki]} B_{[kj]} \iota'_{n_a} D_{[\cdot k]} = \big(B' \operatorname{Diag}(\iota'_{n_a} D)B\big)_{[ij]} \quad \square$$

Appendix B Log determinant for non-cartesian flows

We want to approximate the log determinant in the non-Cartesian case based on the first four terms of its Taylor series expression.

In the following we first recall the intuition of the Martin (1992) approximation of the log-determinant term and then develop the terms required for the fourth-order approximation of the general spatial econometric interaction model. We start by expressing the determinant of as the product of the eigenvalues denoted by $\lambda(A^*)$.

$$\log |A^*| = \log |I_{N^*} - W_F^*| = \log |\Pi_{i=1}^{N^*} 1 - \lambda(W_F)_i|$$

For the next step we require that all eigenvalues of W_F are less than one in magnitude, which allows to remove the absolute value. We then write the log of a products as a sum of logs and replace the logarithm in each term by an infinite Taylor series.

$$\sum_{i=1}^{N^*} \log(1 - \lambda(W_F)_i) = -\sum_{i=1}^{N^*} \sum_{t=1}^{\infty} \frac{\lambda(W_F)_i^t}{t}$$

By interchanging the sums we express the above series in terms of the traces of W_F. With our constraint on the eigenvalues of W_F we are sure that the series converges and we can use the first m terms as an approximation.

$$\sum_{t=1}^{\infty} \sum_{i=1}^{N^*} \frac{\lambda(W_F)_i^t}{t} = \sum_{t=1}^{\infty} \frac{\operatorname{tr}(W_F^t)}{t} \approx \sum_{t=1}^{m} \frac{\operatorname{tr}(W_F^t)}{t}$$

In Section 3.2 we have seen that the traces $\operatorname{tr}(W_F^t)$ may be calculated efficiently from those of $\operatorname{tr}(OW^t)$ and $\operatorname{tr}(DW^t)$, when the model is cartesian. Unfortunately, this is not possible in the non-cartesian version, and we have to expand the terms of W_F^{*t} explicitly to compute the trace values.

$$\operatorname{tr}(W_F^{*t}) = \operatorname{tr}((\rho_d W_d^* + \rho_o W_o^* + \rho_w W_w^*)^t),$$

for each power $t = 1, ..., m$. With a direct approach this would require 120 matrix products for the fourth order approximation, which is unpractical even for moderate sample sizes. We can do much better if we exploit the following properties:

T1 : A well known property of the trace operator is that the trace of a product of matrices is invariant under cyclic permutations of the multiplication order.

T2 : For any $t = 0, 1, 2, ...$, we have $0 = \mathrm{tr}(W_d^{*t} W_o^*) = \mathrm{tr}(W_o^{*t} W_d^*) = \mathrm{tr}(W_d^{*t} W_w^*) = \mathrm{tr}(W_o^{*t} W_w^*)$. This property is demonstrated in the next subsection. It follows from the zero diagonal and the Kronecker product structure of the matrices W_o, W_d an W_w.

T3 : For two square matrices A and B with the same size we have $\mathrm{tr}(AB) = \iota(A \odot B')\iota$, which is a direct consequence of the definition of the trace and the matrix product.

The properties T1 and T2 can be used to avoid the computation of 92 out of 120 values as they are either zero or duplicated. In the following we show the 28 unique trace values that are required for the fourth order approximation:
For $t = 1$, we can use T2 to avoid any computation

$$\mathrm{tr}(W_F^*) = \rho_d \underbrace{\mathrm{tr}(W_d^*)}_{0} + \rho_o \underbrace{\mathrm{tr}(W_o^*)}_{0} + \rho_w \underbrace{\mathrm{tr}(W_w^*)}_{0} = 0$$

For $t = 2$, we can use T2 to conclude that only 3 of the 9 required traces are different from zero.

$$\mathrm{tr}(W_F^{*2}) = \rho_d^2 \,\mathrm{tr}(W_d^{*2}) + \rho_o^2 \,\mathrm{tr}(W_o^{*2}) + \rho_w^2 \,\mathrm{tr}(W_w^{*2})$$

For $t = 3$, we can use T1 and T2 to infer that 12 out of 27 traces are zero and that among the reaming values only seven are unique.

$$\begin{aligned}
\mathrm{tr}(W_F^{*3}) = {} & \rho_d^3 \,\mathrm{tr}(W_d^{*3}) + \rho_o^3 \,\mathrm{tr}(W_o^{*3}) + \rho_w^3 \,\mathrm{tr}(W_w^{*3}) \\
& + 3\rho_d \rho_w^2 \,\mathrm{tr}(W_d^* W_w^{*2}) + 3\rho_o \rho_w^2 \,\mathrm{tr}(W_o^* W_w^{*2}) \\
& + \rho_d \rho_o \rho_w [3 \,\mathrm{tr}(W_d^* W_o^* W_w^*) + 3 \,\mathrm{tr}(W_o^* W_d^* W_w^*)]
\end{aligned}$$

For $t = 4$, we can use T1 and T2 to infer that 16 out of 81 traces are zero and that among the remaining 65 traces only 18 are unique.

$$\begin{aligned}
\mathrm{tr}(W_F^{*4}) = {} & \rho_d^4 \,\mathrm{tr}(W_d^{*4}) + \rho_o^4 \,\mathrm{tr}(W_o^{*4}) + \rho_w^4 \,\mathrm{tr}(W_w^{*4}) \\
& + \rho_d^2 \rho_o^2 [2 \,\mathrm{tr}(W_d^* W_o^* W_d^* W_o^*) + 4 \,\mathrm{tr}(W_d^{*2} W_o^{*2})] \\
& + \rho_d^2 \rho_w^2 [2 \,\mathrm{tr}(W_d^* W_w^* W_d^* W_w^*) + 4 \,\mathrm{tr}(W_d^{*2} W_w^{*2})] \\
& + \rho_o^2 \rho_w^2 [2 \,\mathrm{tr}(W_o^* W_w^* W_o^* W_w^*) + 4 \,\mathrm{tr}(W_o^{*2} W_w^{*2})] \\
& + \rho_d^2 \rho_o \rho_w [4 \,\mathrm{tr}(W_d^* W_o^* W_d^* W_w^*) + 8 \,\mathrm{tr}(W_d^{*2} W_o^* W_w^*)] \\
& + \rho_d \rho_o^2 \rho_w [4 \,\mathrm{tr}(W_d^* W_o^* W_w^* W_o^*) + 8 \,\mathrm{tr}(W_d^* W_o^{*2} W_w^*)] \\
& + \rho_d \rho_o \rho_w^2 [4 \,\mathrm{tr}(W_d^* W_o^* W_w^{*2}) + 4 \,\mathrm{tr}(W_o^* W_d^* W_w^{*2}) + 4 \,\mathrm{tr}(W_d^* W_w^* W_o^* W_w^*)] \\
& + \rho_d \rho_w^3 4 \,\mathrm{tr}(W_d^* W_w^{*3}) + \rho_o \rho_w^3 4 \,\mathrm{tr}(W_o^* W_w^{*3})
\end{aligned}$$

Using T3 we can derive the 28 trace values from only ten matrix products that involve at most two of the weight matrices. An example of this calculation is $\mathrm{tr}(W_d^{*4}) = \iota_{N^*}'(W_d^{*2} \odot W_d^{*2\prime})\iota_{N^*}$. This makes clear that we avoid direct computation of matrix products of third and fourth order that become increasingly dense in comparison to the lower order products. Below are the ten matrices that need to be computed for the fourth-order approximation.

$$\begin{array}{ccccc}
W_d^* & W_o^* & W_w^* & W_d^* W_d^* & W_o^* W_o^* \\
W_w^* W_w^* & W_d^* W_o^* & W_o^* W_d^* & W_d^* W_w^* & W_o^* W_w^*
\end{array}$$

B.1 Proof of T2

We will demonstrate the four statements of in T2 $0 = \text{tr}(W_d^{*t} W_o^*) = \text{tr}(W_o^{*t} W_d^*) = \text{tr}(W_d^{*t} W_w^*) = \text{tr}(W_o^{*t} W_w^*)$ in two steps. At first, we focus on the statements that involve powers of the destination weight matrix W_d^{*t}.

$$
\begin{aligned}
0 &= \text{tr}(W_d^{*t} W_o^*) = \text{tr}((R_\Gamma W_d R_\Gamma)^t W_o R_\Gamma) \\
0 &= \text{tr}(W_d^{*t} W_w^*) = \text{tr}((R_\Gamma W_d R_\Gamma)^t W_w R_\Gamma)
\end{aligned}
\tag{22}
$$

The above development is possible because eliminating or replacing rows and columns by zero has the same effect on the trace value $\text{tr}(S'_\Gamma W_d S_\Gamma) = \text{tr}(R'_\Gamma W_d R_\Gamma)$. From the definition $W_d = I_{n_o} \otimes DW$ it is clear that the first term inside the two traces has the following block-diagonal structure

$$
(R_1 W_d R_1)^t = \begin{bmatrix} (R_1 DW R_1)^t & 0 & \cdots & 0 \\ 0 & (R_2 DW R_2)^t & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & (R_{n_o} DW R_{n_o})^t \end{bmatrix} \tag{23}
$$

where $R_j (n_d \times n_d)$, for $j = 1, ..., n_o$ are the diagonal blogs of the replacement matrix R_Γ. Equation (23) shows that we can concentrate on the diagonal blocks if we compute the traces given (22). If we then partition the matrices W_o and W_w into the same block structure we obtains zero matrices for the diagonal blocks, which makes clear that the traces in (22) are indeed zero.

To demonstrate the remaining two statements

$$
\begin{aligned}
0 &= \text{tr}(W_o^{*t} W_d^*) = \text{tr}((R_\Gamma W_o R_\Gamma)^t W_d R_\Gamma) \\
0 &= \text{tr}(W_o^{*t} W_w^*) = \text{tr}((R_\Gamma W_o R_\Gamma)^t W_w R_\Gamma),
\end{aligned}
\tag{24}
$$

we again partition the matrix $(R_\Gamma W_o R_\Gamma)^t$ into n_o^2 blocks of size $n_d \times n_d$. This leads to a matrix, for which all blocks $D_{ij}(t)$ for $i, j = 1, ..., n_o$ are diagonal matrices, whose entries vary with the powers $t = 1, ..., m$.

$$
(R_\Gamma W_o R_\Gamma)^t = \begin{bmatrix} D_{11}(t) & D_{12}(t) & \cdots & D_{1n_o}(t) \\ D_{21}(t) & D_{22}(t) & \cdots & \cdots \\ & & \ddots & \\ \cdots & \cdots & \ddots & \cdots \\ D_{n_o 1}(t) & \cdots & \cdots & D_{n_o n_o}(t) \end{bmatrix}, \tag{25}
$$

Hence, if we compute the trace of a matrix product that involves $(R_\Gamma W_o R_\Gamma)^t$ we know that values off the main diagonal of the blocks will not play a role. For the matrices W_d and W_w the same partition leads to blocks that have zero diagonal, which confirms the statements in (24).

Appendix C Parameter space

In the following we derive the two results that were used to handle the issue of the feasible parameter space presented in Section 3.3. The first one is formulated as a general theorem and the second one is derived in the specific context of model (7).

C.1 Eigenvalues of a sum of Kronecker products

Proposition: Let $A(n_a \times n_a)$ and $B(n_b \times n_b)$ be two square matrices and denote $M = a(A \otimes I_{n_b}) + b(I_{n_a} \otimes B) + c(A \otimes B)$, where a, b and c are scalar values. The eigenvalue vector $\lambda_M(n_a \cdot n_b \times 1)$ of M and those of the matrices A and B denoted λ_A and λ_B are linked by the relation

$$\lambda_M = a(\lambda_A \otimes \iota_{n_b}) + b(\iota_{n_a} \otimes \lambda_B) + c(\lambda_A \otimes \lambda_B). \tag{26}$$

Proof: By Schurs Lemma we know that for any square matrix X there exists a factorization $X = Q_X U_X Q_X^{-1}$ such that U_X is upper triangular with the main diagonal equal to the eigenvalues vector λ_X of X. Additionally, we may infer the Schur decomposition of a Kronecker product of two matrices $(A \otimes B)$ from the Schur decompositions of A and B.

$$\begin{aligned}(A \otimes B) &= Q_{(A\otimes B)} U_{(A\otimes B)} Q_{(A\otimes B)}^{-1} \\ &= (Q_A \otimes Q_B)(U_A \otimes U_B)(Q_A \otimes Q_B)^{-1} \\ &= (Q_A U_A Q_A^{-1} \otimes Q_B U_B Q_B^{-1})\end{aligned}$$

Next, we may start Schur factorization $M = Q_M U_M Q_M^{-1}$, which is solved for U_M. If we replace Q_M by $(Q_A \otimes Q_B)$ and develop the resulting expression we obtain the following result.

$$\begin{aligned}U_M &= Q_M^{-1} M Q_M \\ &= (Q_A \otimes Q_B)^{-1}[a(A \otimes I_{n_b}) + b(I_{n_a} \otimes B) + c(A \otimes B)](Q_A \otimes Q_B) \\ &= a(Q_A^{-1} A Q_A \otimes Q_B^{-1} I_{n_b} Q_B + b(Q_A^{-1} I_{n_a} Q_A \otimes Q_B^{-1} B Q_B) + c(Q_A^{-1} A Q_A \otimes Q_B^{-1} B Q_B) \\ &= a(U_A \otimes I_{n_b}) + b(I_{n_a} \otimes U_B) + c(U_A \otimes U_B)\end{aligned}$$

Since U_A and U_B are upper triangular matrices we conclude that U_M is upper triangular too, which confirms that we have indeed found a Schur decomposition of M. This triangular structure also allows to conclude that the eigenvalues of U_M are given by equation (26). Since M and U_M have the same spectrum the proof is finished.

C.2 Similarity and the selection operator

In Section 3.3 we rely on the fact that all eigenvalues of W_F^* are contained in the interval $[\lambda_{min}(W_F), \lambda_{max}(W_F)]$. When W_F is symmetric this follows directly from Cauchy's interlacing theorem. Here we show that this argument also hold for the more general form of W_F we have considered in this article.

The properties of $W_F = \rho_d(I_{no} \otimes DW) + \rho_o(OW \otimes I_{nd}) + \rho_w(OW \otimes DW)$ depend on the characteristics of the two node-level neighborhood matrices DW and OW. If these are symmetric, so is WF and Cauchy's theorem applies. When they are row-normalized they are no longer symmetric, but we may still apply Cauchy's theorem if they were symmetric before the row-normalization. Let $\overline{OW}$ and $\overline{DW}$ to symmetric neighborhood matrices. Their row-normalized counterparts are given by

$$DW = D_d \overline{DW} \qquad OW = D_o \overline{OW},$$

where D_d and D_o are diagonal matrices whose entries correspond to the inverse of the row-sums of DW and OW. It is clear that DW is similar to the symmetric matrix $D_d^{1/2} \overline{DW} D_d^{1/2}$ and that

OW is similar to $D_o^{1/2}\overline{OW}D_o^{1/2}$. We may use the same argument to show that W_F is similar a symmetric matrix

$$W_F = D_F\overline{W_F} = D_F^{1/2}(D_F^{1/2}\overline{W_F}D_F^{1/2})D_F^{-1/2},$$

where $D_F = D_o \otimes D_d$ and $\overline{W_F} = \rho_d(I_{no} \otimes \overline{DW}) + \rho_o(\overline{OW} \otimes I_{nd}) + \rho_w(\overline{OW} \otimes \overline{DW})$. The next step is to show that the same holds in the non-cartesian version, where we use the selection operator to obtain $W_F^* = S_\Gamma' D_F \overline{W_F} S_\Gamma$. Since replacing rows with zero before dropping them does not change the result we may write $S_\Gamma' D_F = S_\Gamma' D_F^{1/2} R_\Gamma D_F^{1/2}$. Using this argument and the link $R_\Gamma = S_\Gamma S_\Gamma'$ we obtain

$$W_F^* = (S_\Gamma' D_F^{1/2} S_\Gamma)(S_\Gamma' D_F^{1/2}\overline{W_F}D_F^{1/2} S_\Gamma)(S_\Gamma' D_F^{1/2} S_\Gamma)^{-1}.$$

The above equation clearly shows the similarity of W_F^* to the symmetric matrix $S_\Gamma' D_F^{1/2}\overline{W_F}D_F^{1/2} S_\Gamma$. Since this result does not depend on the specific subset selected by the operator S_Γ we are sure that the row-normalization of OW and DW does not compromise the applicability of Cauchy's interlacing theorem. Hence, we may conclude that the extreme eigenvalues of W_F^* are bounded by those of W_F.