Cluster-wise spatiotemporal panel models for global and local spillovers identification

Caterina Morelli
University of Valle d'Aosta

Paolo Maranzano
University of Milan-Bicocca

Raffaele Mattera
University of Campania "Luigi Vanvitelli"

Philipp Otto
University of Glasgow

Abstract

This paper develops a novel heterogeneous spatiotemporal panel data model that simultaneously accounts for spatial spillovers, dynamic dependence, and group-level heterogeneity. By allowing both spatial interactions and covariate effects to vary across latent clusters of units, the model captures complex diffusion patterns and structural differences that traditional homogeneous models often overlook. The full matrix of direct and indirect effects has been derived, accommodating both intra- and inter-group spatial dependence. To estimate the model, we propose a two-step quasi maximum likelihood (QML) estimation procedure that jointly recovers the heterogeneous parameters and the latent cluster memberships without requiring a predefined partition. The effectiveness of the proposed QML estimation procedure is evaluated through a comprehensive Monte Carlo simulation experiment. The framework is particularly suited to environmental applications, where regions may differ significantly in their socio-economic characteristics and in their capacity to influence or respond to neighboring areas.

Keywords: cluster-wise model, spatiotemporal panel model, heterogeneous spillovers, Quasi-Maximum Likelihood.

1 INTRODUCTION

The growing interconnectedness of the global economy has made it increasingly difficult to view countries, regions, firms or individuals as isolated entities. Economic, technological, and environmental developments now spread rapidly across borders, linking markets and shaping outcomes in ways that traditional, independent-unit models struggle to capture. Such interdependence may stem from direct strategic interactions—when the decisions of one agent influence others—or from broader, shared forces, such as global technological change, shifts in regulation, or common macroeconomic shocks. At the same time, the degree and nature of these interconnections often depend on the intrinsic characteristics of the units themselves—such as their economic strength, institutional capacity, or technological readiness—which determine not only how strongly they can influence others but also how easily they can adapt to or emulate external changes.

To account for these connections, two major modeling traditions have emerged in the econometric literature. The first, the spatial approach, assumes that the behavior of one unit is directly affected by that of its neighbors, whether through geographical proximity, trade links, or economic integration. The strength of these relationships is often captured using a spatial weighting matrix that encodes how strongly each unit is connected to others. The second, the factor-based approach, models dependence as the result of unobserved common components—latent global shocks or trends—that influence all units, albeit with varying intensity. While these two frameworks were initially developed in parallel, recent advances increasingly seek to bridge them, offering unified ways to capture both local and global sources of interdependence (Koopman et al., 2021; Chen et al., 2022; Su et al., 2023). The spatial sciences, particularly geography and regional economics, provide a solid conceptual foundation for understanding spatial dependencies. A central insight in this literature is encapsulated in the well-known *First Law of Geography* (Tobler, 1970), which states that "everything is related to everything else, but near things are more related than distant things". This principle underlies the pervasive presence of spatial autocorrelation—the tendency for nearby units to exhibit similar values due to direct or indirect interactions across space. These interactions often manifest as spatial spillovers, that is, effects in one unit that influence outcomes in geographically or economically connected areas (Elhorst et al., 2021; Elhorst, 2024). Such spillovers are not limited to geographic contiguity: they may arise from structural connections, such as trade, labor mobility, or vertical industrial integration. For instance, a shock in one region or sector may propagate to others via trade links or supply chain dependencies.

A second, equally important spatial principle is the *Second Law of Geography*, also known as spatial heterogeneity (Goodchild, 2004; Zhu and Turner, 2022; Zhu et al., 2018). This law emphasizes that the relationships between variables are not constant across space; instead, they vary with location due to differences in economic structure, social institutions, infrastructure, or policy environments. Such heterogeneity violates the assumption of identically distributed relationships across observations and renders classical regression models inadequate. In this setting, statistical models that fail to account for spatial variation may suffer from misspecification, resulting in biased or inconsistent parameter estimates (LeSage and Pace, 2009; Anselin and Arribas-Bel, 2013).

The importance of modeling both spatial dependence and heterogeneity has spurred the development of spatially-structured clustering techniques. These approaches, including recent extensions of clusterwise regression models (Sugasawa and Murakami, 2021; Lee et al., 2017; Billé et al., 2017), seek to identify groups of spatial units that not only exhibit internally similar behavior but also account for the spatial organization of the data. For example, penalized spatial clustering models integrate neighborhood structure—such as that encoded by Potts-type priors—to encourage geographically cohesive clusters, while simultaneously estimating group-specific regression coefficients. In doing so, they offer a unified framework to tackle spatial dependence and spatial heterogeneity together (R. et al., 2025; Boccaletti et al., 2024; Cerqueti et al., 2025).

Yet, many empirical applications continue to treat spatial units as homogeneous, ignoring the substantial differences in structural and behavioral characteristics that exist across regions, cities, or countries. These differences can relate to human capital, technological capacity, governance quality, and environmental exposure. For instance, the same environmental regulation may elicit different responses across regions, depending on local industrial composition or policy capacity. A technologically advanced region may quickly adopt green innovations, while a neighboring, less developed area may struggle to do so, leading to asymmetric spatial externalities.

These observations point to two core mechanisms driving spatial outcomes: (i) spatial dependence, where a unit's outcome is influenced by neighboring units through interaction or diffusion; and (ii) spatial heterogeneity, where units differ in their structural responses due to contextual features. Both mechanisms are essential to understanding complex spatial phenomena, particularly when the underlying processes evolve over space and time.

In this paper, we propose a novel cluster-wise spatial dynamic panel model that captures these dimensions in a unified framework. Our approach integrates (i) a spatial dynamic panel structure that models spillovers through spatial connectivity matrices (Elhorst, 2010; Elhorst et al., 2021); and (ii) a clusterwise regression design in which parameters vary across latent groups (Bonhomme and Manresa, 2015).

Crucially, we allow for spatial dependence to differ both within and across clusters, thus accommodating heterogeneity in both marginal effects and spatial transmission mechanisms.

The proposed model is flexible and interpretable. It assigns each region to a cluster with group-specific dynamics, while allowing for rich spatial interconnections across all units. Spatial spillovers are weighted not only by proximity but also by the identity of the sending and receiving groups, thereby modeling both symmetric and asymmetric effects. This design is especially relevant in contexts such as greenhouse gas (GHG) emissions, where both regional disparities and spatial externalities are critical. By embedding cluster-specific behavior into the spatial dynamics, our model offers a powerful tool to analyze spatially heterogeneous processes with interdependent units.

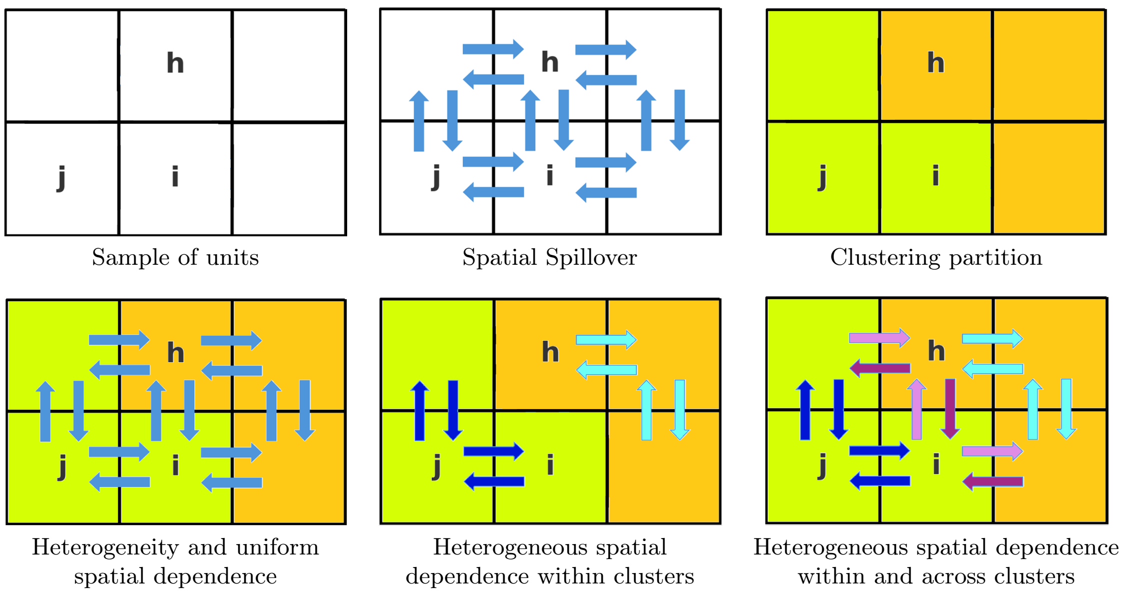


Figure 1: Conceptual representation of spatial dependence and heterogeneity by clusters.

Figure 1 provides a step-by-step conceptual illustration of the key mechanisms captured by our model. The top-left panel displays a generic spatial grid of units, labeled i, j, and h, which serves as the spatial reference structure. The top-center panel introduces the idea of spatial spillovers: here, blue arrows represent the bilateral influence that neighboring regions exert on one another, irrespective of their individual characteristics. The top-right panel adds a partitioning structure, where each unit belongs to a specific cluster (indicated by different background colors), capturing the notion of group-level heterogeneity.

The bottom-left panel combines these two aspects by illustrating uniform spatial dependence across heterogeneous clusters. Spatial spillovers (same arrows for all units) are present, but the internal structure of each group (yellow vs. green background) is heterogeneous. The bottom-center panel introduces cluster-specific spatial dependence: spillovers still occur only within each cluster, but the intensity and pattern differ across clusters, as shown by the variation in arrow color and direction. This reflects the idea that regions within the same group may interact more strongly with each other than with regions outside their group.

Finally, the bottom-right panel captures the most general and flexible case: heterogeneous spatial dependence both within and across clusters. Arrows of different colors and directions represent varying spatial influence across unit pairs, with the structure depending not only on spatial proximity but also on the cluster identity of both the sending and receiving units. This is the most general setting captured by our proposed model, which is designed to flexibly accommodate both spatial dynamics and structured group heterogeneity.

In summary, our contribution lies in the development of a flexible spatial dynamic panel model that simultaneously captures interregional dependencies and cluster-based heterogeneity in both covariate effects and spatial spillovers. We provide a quasi–maximum likelihood (QML) estimation procedure to estimate the heterogeneous spatial and temporal dependence parameters, conditional on a given clustering partition. Furthermore, we design a comprehensive simulation experiment to assess the performance of the proposed estimation method in recovering the true parameters and in accurately identifying the direct and indirect effects implied by the model, considering different model specification and heterogeneity settings. In addition, we introduce a two-step procedure that enables the joint identification of model parameters and endogenous cluster membership, thereby eliminating the need to predefine a partition structure.

Looking ahead, we intend to apply this framework to real-world environmental data to demonstrate its practical relevance. Prior studies (Morelli et al., 2025) have revealed strong evidence of spatial clustering in GHG emission patterns and heterogeneous dynamics across regions. These findings validate our modeling approach and motivate further investigation using our new framework. By integrating spatial dynamics with group-specific heterogeneity, our work contributes to the growing literature on the spatiotemporal modeling of economic and environmental phenomena, offering valuable insights for more effective and targeted regional policy design.

The remainder of the paper is organized as follows. Section 2 introduces the proposed model and outlines its main features. Section 3 discusses the possible structures of parameter heterogeneity. In Section 4, we derive the expressions for the direct and indirect effects, distinguishing between cluster-specific and cross-cluster components.

Section 5 presents the estimation procedure, while Section 6 describes the simulation experiment and discusses the corresponding results. Finally, Section 7 concludes the paper.

2 METHODOLOGY

Our model incorporates both spatial dependence and group heterogeneity in a dynamic panel data framework. Suppose we have a panel data, with N units for T time stamp, thus y_{it} is the dependent variable for unit i at time t, $\mathbf{x}_{it}$ is the vector of k explanatory variables for the same unit and time. Moreover, let $\mathcal{P} = (\mathcal{C}_1, \dots, \mathcal{C}_G)$ being the partition in G number of clusters. Each unit in our sample belongs to one and only one cluster for the entire time period.

Also, suppose we have spatial information regarding each unit i with respect to the other units j, which can be interpreted as a measure of spatial proximity or boundaries, and is quantified and defined as w_{ij}. The spatial dependencies in the SDM are captured using a spatial weight matrix W, which is an $N \times N$ matrix where N is the number of spatial units. Each element w_{ij} in W represents the spatial proximity or influence between spatial unit i and unit j. By convention, $w_{ii} = 0$, as no unit is considered its own neighbor Elhorst (2010).

To ensure model stability, the spatial weight matrix W must satisfy several assumptions LeSage (2008); LeSage and Pace (2009); Elhorst (2010). First, W should be a non-negative matrix of known constants. Additionally, W is often row-normalized, meaning the elements in each row of W sum to one, which ensures that the influence of neighboring units is proportionally scaled. For symmetric W, the eigenvalues must lie within the interval $(1/\omega_{\max}, 1)$, where $\omega_{\max}$ is the largest real eigenvalue of W. If W is not normalized, the interval instead depends on $\omega_{\min}$ and $\omega_{\max}$, the smallest and largest real eigenvalues of W, respectively. Finally, the matrices $I - \rho W$ and $I - \lambda W$ must be non-singular, where I is the identity matrix of size $N \times N$, and ρ and λ are spatial autoregressive coefficients.

Our model can be defined as follows:

$$\begin{aligned} y_{it} = & \sum_{j=1}^{N}\sum_{f=1}^{G} \rho_{gf} w_{ij} y_{jt} I(j \in \mathcal{C}_f) + \tau_g y_{i,t-1} + \sum_{j=1}^{N}\sum_{f=1}^{G} \eta_{gf} w_{ij} y_{j,t-1} I(j \in \mathcal{C}_f) \\ & + \mathbf{x}'_{it}\boldsymbol{\beta}_g + \sum_{j=1}^{N}\sum_{f=1}^{G} w_{ij}\mathbf{x}'_{jt}\boldsymbol{\delta}_{gf} I(j \in \mathcal{C}_f) + \epsilon_{it}, \quad i \in \mathcal{C}_g \end{aligned} \tag{1}$$

The parameter ρ_{gf} captures the contemporaneous spatial spillover from units in cluster f to units in cluster g, weighted by the spatial proximity defined through w_{ij}. When $g = f$, ρ_{gg} measures within-cluster endogenous spatial dependence, reflecting how the current outcomes of spatially neighboring units within the same cluster influence each other. For simplicity, in this case the notation will be simplified as $\rho_g := \rho_{gg}$. When $g \neq f$, ρ_{gf} quantifies cross-cluster spatial spillovers, that is, the extent to which units in cluster f contemporaneously affect the outcomes of neighbor units belonging to cluster g.

The parameter τ_g represents the group-specific temporal dependence. It captures the autoregressive effect of the lagged dependent variable $y_{i,t-1}$ on the current outcome y_{it} for units in cluster g. Thus, the strength of this persistence may vary by group.

The parameter η_{gf} accounts for lagged spatial dependence across clusters. Specifically, it measures the effect of past outcomes $y_{j,t-1}$ of units in cluster f on the current outcomes of units in cluster g, again weighted by the spatial structure w_{ij}. When $g = f$, this represents within-cluster dynamic spatial spillovers, similarly to the contemporaneous spatial dependence, the notation will be simplified as $\eta_g := \eta_{gg}$. When $g \neq f$, it captures delayed cross-cluster spatial interactions.

The vector $\boldsymbol{\beta}_g$ contains the group-specific coefficients associated with the set of k covariates $\mathbf{x}_{it}$. These coefficients represent the direct marginal effects of explanatory variables on the outcome for units in cluster g, allowing for heterogeneous covariate effects across groups.

The parameter vector $\boldsymbol{\delta}_{gf}$ captures the spatial spillovers of k covariates. It measures the impact of covariates $\mathbf{x}_{jt}$ from units in cluster f on the outcome y_{it} of units in cluster g, with the effect weighted by the spatial link w_{ij}. When $g = f$, this term reflects within-cluster spatial externalities of covariates. When $g \neq f$, it accounts for cross-cluster spatial spillovers of covariate effects.

Finally, the error term ϵ_{it} captures unobserved, idiosyncratic disturbances, assuming that $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$.

Altogether, the model is designed to flexibly capture both temporal dynamics and spatial dependencies in the data while accounting for heterogeneity in behavior across different clusters. The structure permits not only endogenous and exogenous spatial effects, but also allows these effects to vary depending on whether interactions occur within the same group or across different groups.

The specification in Equation (1) represents the most general form of our proposed model, encompassing all possible sources of spatial dependence, temporal dynamics, and their cross-group interactions. However, as also emphasized in the spatial econometric literature (Anselin et al., 2008; Elhorst, 2010), such a general model may suffer from identification issues when all parameters are included simultaneously. In practice, it is common to work with restricted versions of this model, where specific parameters are set to zero based on theoretical assumptions or empirical considerations.

Therefore, in the remainder of this paper, we treat the full model (1) primarily as a theoretical benchmark, used to illustrate the conceptual framework and to describe the most general case. Nonetheless, our estimation and empirical strategy will focus on more tractable, restricted versions. A comprehensive overview of these nested specifications—including the Spatial Autoregressive Model (SAR), Spatial Durbin Model (SDM), and other dynamic extensions—is provided in Elhorst (2001), where all potential restrictions are discussed systematically, up to the case of the standard linear regression model.

2.1 RESTRICTION TO SPATIAL DYNAMIC PANEL MODEL

If we remove the heterogeneity in the model—namely, we assume that all units belong to a single cluster and are governed by a common set of parameters—we recover the well-known Spatial Durbin Dynamic model, as presented in its general form by Elhorst (2001). The model is expressed as:

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + \tau y_{i,t-1} + \eta \sum_{j=1}^{N} w_{ij} y_{j,t-1} + \mathbf{x}_{it}' \boldsymbol{\beta} + \sum_{j=1}^{N} w_{ij} \mathbf{x}_{jt}' \boldsymbol{\delta} + \epsilon_{it} \tag{2}$$

Equation 2 can be viewed as a special case of the more general heterogeneous model previously described, obtained by imposing the restriction that all parameters $\rho, \tau, \eta, \boldsymbol{\beta}, \boldsymbol{\delta}$ are common to all units. Several restricted versions of this model can be derived by constraining one or more of these parameters, as discussed in LeSage and Pace (2009); Elhorst (2010). In this specification, the effects of the explanatory variables and their spatial lags, as well as the spatial and temporal dependence of the dependent variable, are assumed to be homogeneous across space and time. While this allows the model to capture average spatial spillovers and dynamic effects, it does not account for potential heterogeneity in how units respond to covariates or interact with their neighbors—an important limitation when units are diverse or grouped by underlying structural differences.

2.2 RESTRICTION TO HETEROGENEOUS CLUSTER MODEL

If we restrict the model in Equation 1 by removing both spatial and temporal dependence, we obtain a cluster-wise heterogeneous model as proposed by Bonhomme and Manresa (2015), which takes the form:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_g + \epsilon_{it}, \quad i \in \mathcal{C}_g \tag{3}$$

In this specification, units are assumed to be independent of one another, implying that neither spatial proximity nor common economic characteristics generate interaction effects across units. Each unit evolves independently, unaffected by its neighbors. However, the model retains heterogeneity through the clustering structure: for each unit i belonging to a given cluster $\mathcal{C}_g$, a specific set of coefficients $\boldsymbol{\beta}_g$ governs the relationship between the covariates and the outcome. This allows the model to capture structural differences across groups while abstracting from spatial and dynamic interdependence.

2.3 RESTRICTION TO POOLED PANEL MODEL

If we further restrict the model in Equation 1 by removing spatial and temporal dependence as well as the cluster-based heterogeneity, we obtain the standard pooled regression model for panel data. This specification assumes that all units share a common set of parameters and evolve independently over time and space. The resulting model is given by:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \epsilon_{it}, \quad \forall i = 1, \ldots, N;\ \forall t = 1, \ldots, T \tag{4}$$

This pooled model imposes homogeneity across all cross-sectional units and time periods, thereby ignoring both interaction effects and structural heterogeneity. While computationally simple, such a specification may be too restrictive in settings where spatial spillovers, temporal dynamics, or unit-specific behaviors play a substantive role.

2.4 MATRIX REPRESENTATION OF THE MODEL

To express the model in compact matrix notation, let us define the key components across all N units and T time periods. Denote $Y_t = (y_{1t}, \ldots, y_{Nt})'$ as the $N \times 1$ vector of dependent variables at time t, and $X_t = (\mathbf{x}_{1t}', \ldots, \mathbf{x}_{Nt}')'$ as the $N \times k$ matrix of covariates. Let $W \in \mathbb{R}^{N \times N}$ be the spatial weight matrix, and $P \in \{0,1\}^{G \times N}$ the group membership matrix, where each column of P has exactly one entry equal to 1, indicating the group membership of each unit.

The group-level parameters are collected as follows:

- $\boldsymbol{\tau} \in \mathbb{R}^G$: vector of group-specific temporal autoregressive coefficients;
- $\boldsymbol{\beta} \in \mathbb{R}^{G\times k}$: matrix of group-specific covariate effects;
- $\boldsymbol{\rho}, \boldsymbol{\eta}, \boldsymbol{\delta} \in \mathbb{R}^{G\times G}$: matrices of group-to-group spatial effects for the contemporaneous dependent variable, its lag, and the covariates, respectively.

Using this notation, the full model in Equation 1 can be rewritten as:

$$\begin{aligned} Y_t = & \left(P^\top \boldsymbol{\rho} P \circ W\right) Y_t + \mathrm{diag}(P^\top \boldsymbol{\tau}) Y_{t-1} + \left(P^\top \boldsymbol{\eta} P \circ W\right) Y_{t-1} \\ & + \mathrm{diag}_k(P^\top \boldsymbol{\beta}) X_t + \left(P^\top \boldsymbol{\delta} P \circ W\right) X_t + \epsilon_t, \end{aligned} \tag{5}$$

where:

- $\epsilon_t \sim \mathcal{N}(0, \sigma^2 I_N)$ is the vector of idiosyncratic errors;
- $\mathrm{diag}(P^\top \boldsymbol{\tau}) \in \mathbb{R}^{N\times N}$ is a diagonal matrix assigning τ_g to units in group g;
- $\mathrm{diag}_k(P^\top \boldsymbol{\beta})$ represents a block-diagonal matrix of dimension $N \times k$, where each row i takes the corresponding $\boldsymbol{\beta}_g$ according to unit i's group;
- The term $P^\top \boldsymbol{\rho} P \circ W$ is the element-wise (Hadamard) product between the matrix assigning each dyadic pair (i, j) the appropriate ρ_{gf} based on their group membership and the spatial weight matrix W. The same applies to the $\boldsymbol{\eta}$ and $\boldsymbol{\delta}$ terms.

This compact matrix formulation highlights how group-level heterogeneity enters the spatial and temporal dynamics of the model. Each unit's outcome depends on its own lag, the outcomes and covariates of its spatial neighbors, and all corresponding effects vary by the group of both the source and receiving units. This notation will be used in the following sections to describe the structure of heterogeneity and to derive the expressions for direct and indirect effects.

3 TYPES AND STRUCTURE OF HETEROGENEITY

The proposed model incorporates two distinct dimensions of heterogeneity, which can be classified according to the structure of the model parameters. Let us define the complete set of model parameters as Θ, which can be partitioned into two subsets:

- $\boldsymbol{\varphi}_g = \{\boldsymbol{\beta}_g, \tau_g\}$: parameters that depend only on the cluster membership of unit i, with a single index g, corresponding to the group $\mathcal{C}_g$ to which unit i belongs.

- $\boldsymbol{\psi}_{gf} = \{\rho_{gf}, \eta_{gf}, \boldsymbol{\delta}_{gf}\}$: parameters that capture spatial spillover effects between units in cluster g and units in cluster f, hence indexed by both origin and destination groups.

The heterogeneity in $\boldsymbol{\varphi}_g$ is standard and has been studied extensively in the literature, for example in Bonhomme and Manresa (2015). In this case, the heterogeneity is cluster-wise, meaning that there are G distinct sets of parameters, one for each cluster. These parameters affect only the units belonging to the same group and are not influenced by spatial relationships with other units.

On the other hand, the heterogeneity in $\boldsymbol{\psi}_{gf}$ is more complex. Since these parameters depend on both the source and target cluster, there can potentially be $G \times G$ different values for each element in $\boldsymbol{\psi}_{gf}$. To better understand this structure, consider the spatial weight matrix W, which defines spatial proximity across the N units in the sample. Assuming that units are partitioned into G clusters, we can reorganize the matrix W according to the partition $\mathcal{P}^G$, such that W consists of G^2 block matrices: G square blocks on the diagonal and $G(G-1)$ rectangular off-diagonal blocks.

Each diagonal block W_{gg} captures the spatial proximity between units within the same cluster $\mathcal{C}_g$, whereas each off-diagonal block W_{gf} (with $g \neq f$) captures spatial relationships between units in different clusters. In the most general case, each submatrix we can associate a different effect, thus estimate different parameters, but specifically we can consider different cases for submatrices on the diagonal and different cases for matrices outside the diagonal, then combine the possible forms of heterogeneity.

Considering the diagonal elements W_{gg}, we can identify three main cases for within-cluster spillovers:

1. **Null case:** $\boldsymbol{\psi}_{gg} = 0$ for all g, indicating no spatial spillovers within any cluster.
2. **Homogeneous case:** $\boldsymbol{\psi}_{gg} = \psi$, constant across all clusters. This assumes that within-cluster spillovers exist but are uniform across groups.
3. **Heterogeneous case:** $\boldsymbol{\psi}_{gg} = \psi_g$, allowing each cluster to have its own level of within-cluster spatial dependence.

These cases are illustrated below, following the matrix structure shown in the diagram:

$$\psi_{gg} = \begin{bmatrix} 0 & \cdot & \cdot \\ \cdot & 0 & \cdot \\ \cdot & \cdot & 0 \end{bmatrix} \quad \text{(Null)}$$

$$\psi_{gg} = \begin{bmatrix} \psi & \cdot & \cdot \\ \cdot & \psi & \cdot \\ \cdot & \cdot & \psi \end{bmatrix} \quad \text{(Homogeneous)}$$

$$\psi_{gg} = \begin{bmatrix} \psi_1 & \cdot & \cdot \\ \cdot & \psi_2 & \cdot \\ \cdot & \cdot & \psi_3 \end{bmatrix} \quad \text{(Heterogeneous)}$$

Regarding the off-diagonal blocks W_{gf} referring to the spatial spillovers across clusters, we identify the following configurations:

1. **Null case:** $\boldsymbol{\psi}_{gf} = 0$ for all $g \neq f$, implying no cross-cluster spillovers.
2. **Homogeneous case:** $\boldsymbol{\psi}_{gf} = \bar{\psi}$, the same for all cross-cluster interactions, regardless of origin and destination group.
3. **Heterogeneous symmetric case:** $\boldsymbol{\psi}_{gf} = \boldsymbol{\psi}_{fg}$, allowing for variation in cross-cluster spillovers but assuming symmetry in the direction of influence.
4. **Heterogeneous asymmetric case:** $\boldsymbol{\psi}_{gf} \neq \boldsymbol{\psi}_{fg}$, where the strength of the spatial influence from cluster g to f differs from the reverse direction.

These configurations are illustrated below using stylized matrices, consistent with the structure provided:

$$\psi_{gf} = \begin{bmatrix} \cdot & 0 & 0 \\ 0 & \cdot & 0 \\ 0 & 0 & \cdot \end{bmatrix} \quad \text{(Null)}$$

$$\psi_{gf} = \begin{bmatrix} \cdot & \psi & \psi \\ \psi & \cdot & \psi \\ \psi & \psi & \cdot \end{bmatrix} \quad \text{(Homogeneous)}$$

$$\psi_{gf} = \begin{bmatrix} \cdot & \psi_{12} & \psi_{13} \\ \psi_{12} & \cdot & \psi_{23} \\ \psi_{13} & \psi_{23} & \cdot \end{bmatrix} \quad \text{(Heterogeneous Symmetric)}$$

$$\psi_{gf} = \begin{bmatrix} \cdot & \psi_{12} & \psi_{13} \\ \psi_{21} & \cdot & \psi_{23} \\ \psi_{31} & \psi_{32} & \cdot \end{bmatrix} \quad \text{(Heterogeneous Asymmetric)}$$

The flexibility of this framework allows for a wide range of combinations. For instance, one may consider a model with heterogeneous within-cluster spillovers (ψ_g) but null cross-cluster effects ($\psi_{gf} = 0$), which captures localized spatial dynamics without inter-cluster influence. Alternatively, one could assume homogeneous within-cluster dependence and heterogeneous asymmetric cross-cluster effects, which would model consistent internal dynamics but directional interactions across groups. At the extreme, the fully null case ($\psi_{gf} = 0$ and $\psi_{gg} = 0$) corresponds to a spatially independent model, where no spatial spillovers exist at all.

This layered approach to modeling heterogeneity enables us to explore how structural groupings (e.g., geographic regions, economic typologies) modulate the strength and direction of spatial interactions across space and time.

4 DIRECT AND INDIRECT EFFECTS

To quantify the impact of covariates in the presence of spatial and temporal dependence with group-level heterogeneity, we derive the matrix of partial derivatives of the dependent variable Y_t with respect to the covariates X_t. Consider the reduced form of the model:

$$\frac{\partial Y_t}{\partial X_t} = \left(I_N - \operatorname{diag}(P^\top \boldsymbol{\tau}) - P^\top \boldsymbol{\rho} PW - P^\top \boldsymbol{\eta} PW\right)^{-1} \left[\operatorname{diag}(P^\top \boldsymbol{\beta}) + P^\top \boldsymbol{\delta} PW\right],$$

where:

- $\boldsymbol{\tau} \in \mathbb{R}^G$: group-specific temporal autoregressive coefficients,
- $\boldsymbol{\rho}, \boldsymbol{\eta}, \boldsymbol{\delta} \in \mathbb{R}^{G\times G}$: group-to-group spatial interaction matrices for the contemporaneous dependent variable, its lag, and covariates respectively,
- $P \in \{0,1\}^{G\times N}$: group membership matrix,
- $W \in \mathbb{R}^{N\times N}$: spatial weights matrix.

The term $\left(I_N - \operatorname{diag}(P^\top \boldsymbol{\tau}) - P^\top \boldsymbol{\rho} PW - P^\top \boldsymbol{\eta} PW\right)^{-1}$ acts as a generalized spatial-temporal multiplier, capturing the propagation of effects over both space and time. In contrast to the standard spatial Durbin model, here the feedback strength varies across units depending on their group membership and the spatial structure of inter-group dependencies.

The second term, $\left[\operatorname{diag}(P^\top \boldsymbol{\beta}) + P^\top \boldsymbol{\delta} PW\right]$, captures the instantaneous effect of covariates and their spatial lags, adjusted according to the group of the target unit (via $\boldsymbol{\beta}$) and the group of the neighboring units (via $\boldsymbol{\delta}$).

As in LeSage and Pace (2009), the matrix $\partial Y_t / \partial X_t$ can be decomposed into:

- **Direct effects**: the average of the diagonal entries, which measure how a unit's own covariate affects its own outcome, accounting for spatial and temporal feedback.
- **Indirect effects (spillovers)**: the average of the off-diagonal entries, representing the impact of a unit's covariate on the outcomes of other units through spatial interactions.

Importantly, due to the presence of the temporal component $\boldsymbol{\tau}$ and the lagged spatial component $\boldsymbol{\eta}$, the matrix of partial effects describes not only contemporaneous responses but also their propagation through time. This dynamic propagation is embedded in the structure of the multiplier matrix. The group-specific structure of the parameters allows for significant flexibility, enabling different spatial intensities and feedback mechanisms within and across groups. This leads to a rich set of patterns for both direct and indirect effects, which can be summarized by block patterns in the derivatives matrix. The diagonal blocks represent intra-group effects, while the off-diagonal blocks capture inter-group spillovers.

The full matrix of partial derivatives $\partial Y_t / \partial X_t$ has the following extended structure:

$$
\begin{aligned}
\frac{\partial Y}{\partial X} &= \left[\frac{\partial Y}{\partial x_1}\frac{\partial Y}{\partial x_2}\cdots\frac{\partial Y}{\partial x_N}\right] = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_N} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_N}{\partial x_1} & \frac{\partial y_N}{\partial x_2} & \cdots & \frac{\partial y_N}{\partial x_N} \end{bmatrix} \\
&= \begin{bmatrix} I-\tau_1 I-\rho_1 W_1-\eta_1 W_1 & -\rho_{12}W_{12}-\eta_{12}W_{12} & \cdots & -\rho_{1G}W_{1G}-\eta_{1G}W_{1G} \\ -\rho_{21}W_{21}-\eta_{21}W_{21} & I-\tau_2 I-\rho_2 W_2-\eta_2 W_2 & \cdots & -\rho_{2G}W_{2G}-\eta_{2G}W_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ -\rho_{G1}W_{G1}-\eta_{G1}W_{G1} & -\rho_{G2}W_{G2}-\eta_{G2}W_{G2} & \cdots & I-\tau_G I-\rho_G W_G-\eta_G W_G \end{bmatrix}^{-1} \\
&\quad\times \begin{bmatrix} \beta_1 I+\delta_{11}W_{11} & \delta_{12}W_{12} & \cdots & \delta_{1G}W_{1G} \\ \delta_{21}W_{21} & \beta_2 I+\delta_{22}W_{22} & \cdots & \delta_{2G}W_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{G1}W_{G1} & \delta_{G2}W_{G2} & \cdots & \beta_G I+\delta_{GG}W_{GG} \end{bmatrix}
\end{aligned}
\tag{6}
$$

This matrix is block-partitioned, where each block corresponds to the effect of the covariates from one group on the outcomes of another. The diagonal blocks represent the direct effects (including spatial feedback), while the off-diagonal blocks represent the indirect (spillover) effects across groups. These spillovers are heterogeneous and direction-specific, depending on both origin and destination groups via the corresponding ρ_{gf} and δ_{gf} parameters.

As in LeSage and Pace (2009) and Elhorst (2010), we summarize the overall direct effect as the average of the diagonal elements of the full matrix $\partial Y_t/\partial X_t$, while the overall indirect effect is the average of the off-diagonal elements. However, given the group-specific nature of our model, these effects can be decomposed more finely to reflect intra- and inter-group heterogeneity:

- **Direct effects within clusters:** For each group g, the direct effect is computed as the average of the diagonal elements of the corresponding diagonal block in the derivative matrix (i.e., the g-th block on the main diagonal). This captures the impact of a region's own covariates on its own outcomes within the same cluster, including feedback through space and time.
- **Indirect effects within clusters:** Still within the diagonal blocks, the off-diagonal elements represent the intra-group spillovers—i.e., how the covariates of one region within a cluster influence the outcomes of other regions in the same cluster. Their average provides a measure of within-cluster indirect effects.
- **Indirect effects across clusters:** The off-diagonal blocks in the matrix correspond to spatial spillovers between different clusters. Averaging all elements in these off-diagonal blocks yields the cross-cluster indirect effects, capturing how a region's covariates influence the outcomes of regions in other clusters through the spatial network.

This decomposition enables a nuanced interpretation of the effects and reveals how the influence of covariates propagates not only locally but also across group boundaries, depending on both spatial proximity and structural similarity captured by the clustering.

5 ESTIMATION PROCEDURE

In this section, we present the full estimation procedure for the proposed heterogeneous spatiotemporal model. In the first part, we describe a two-step procedure that jointly estimates all model parameters and identifies the group membership of each unit. This approach requires no prior assumptions about the cluster structure or knowledge of the underlying parameter values, allowing for a fully data-driven estimation of both heterogeneity and spatial dynamics.

In the second part, we focus on the estimation of the spatiotemporal dependence parameters and covariate effects, assuming the group partition is known or given. Here, we propose a quasi maximum likelihood (QML) estimation method, which extends existing approaches commonly used in spatial dynamic panel models. Unlike standard methods, which typically assume parameter homogeneity, our procedure explicitly incorporates cluster-based heterogeneity in both spatial and temporal components.

5.1 TWO STEP PROCEDURE

The parameters to be estimated in our model include both the regression coefficients and the group membership configuration. Formally, we denote the full set of regression parameters as

$$\boldsymbol{\Theta} = (\boldsymbol{\rho}, \boldsymbol{\tau}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma),$$

and the group structure as $\mathcal{P} = \{\mathcal{C}_1, \ldots, \mathcal{C}_G\}$, where each $\mathcal{C}_g$ denotes a cluster and every unit belongs to one and only one cluster. Due to the complexity arising from the unknown group membership and the spatial dependence, we adopt a two-step estimation approach based on maximum likelihood, following the strategy proposed in Sugasawa and Murakami (2021).

1. **Step 1: Estimation of model parameters given a fixed partition**
   In the first step, we assume that the group membership $\mathcal{P}$ is known and fixed. The initial partition may come from cluster analysis techniques on a set of variables or from the economic literature. Given this configuration, we estimate the structural parameters $\boldsymbol{\Theta}$ by maximizing the log-likelihood function. Specifically, the estimator is defined as:

   $$\widehat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} \mathcal{L}(Y|X, W, \boldsymbol{\Theta}, \mathcal{P}^G)$$

   where $\mathcal{L}(\cdot)$ denotes the conditional density function of the sample i, conditional on covariates, spatial weights, and cluster assignment. Since the likelihood function involves spatial dependence across observations, we cannot assume independence between units, and standard MLE must be adapted. Therefore, we follow the QML framework, similar to Yu et al. (2008); Lee and Yu (2010a,b) and as implemented by Simonovska (2025). In the following subsection we provide an extension of their approach to the case with multiple spatial parameters.
2. **Step 2: Estimation of group membership given model parameters**
   In the second step, we consider the reverse problem: assume the model parameters $\boldsymbol{\Theta}$ are known (e.g., from Step 1), and update the group membership

for each unit. The assignment of unit i to a group g_i is determined by maximizing its local log-likelihood, taking into account not only the unit's own outcome but also the outcomes of its neighbors. The estimation rule is defined as:

$$\widehat{g}_i = \arg \max_{g_i \in \{1,\dots,G\}} \left\{ \log f_i(y_i \,|\, \mathbf{x}_{i\cdot}, W, \boldsymbol{\Theta}, g_i) + \sum_{j=1}^{N} \log f_j(\cdot)\, \mathbb{I}(w_{ij} > 0) \right\}.$$

This step reflects the spatially dependent nature of the data: units that are geographically or economically close may influence one another, and therefore their group assignments should not be estimated in isolation. The second term in the maximand explicitly incorporates the log-likelihood contributions of neighboring units (i.e., those for which $w_{ij} > 0$), thereby acknowledging the violation of conditional independence across spatial units. Intuitively, this formulation introduces local spatial smoothing in the group assignment process, reducing the likelihood of assigning a unit to a cluster that is inconsistent with the behavior of its neighbors. Compared to previous two-step methodologies in cluster-wise regression, which typically consider each unit's likelihood independently—thus neglecting cross-sectional dependence—our approach explicitly accounts for the influence of neighboring units. In particular, the likelihood contribution of unit i depends not only on its own characteristics and group assignment, but also on the likelihoods and cluster memberships of spatially connected units. Depending on the model size and the number of clusters, this step can be implemented through exhaustive enumeration (in small settings), or via more scalable methods such as a classification expectation-maximization (CEM) algorithm or a greedy assignment update, alternating with Step 1 until convergence.

This two-step procedure can be iterated until the estimated group membership and parameter estimates converge. Convergence is typically assessed by tracking changes in the log-likelihood or in the clustering assignments across iterations. This approach enables flexible modeling of group-level heterogeneity while fully accounting for spatial dependence and feedback effects in the data.

This two-step procedure can be iterated until the estimated group membership and parameter estimates converge. Convergence is typically assessed by monitoring changes in the log-likelihood or in the clustering assignments across successive iterations. This iterative approach enables flexible modeling of group-level heterogeneity while fully accounting for spatial dependence and feedback effects present in the data.

Formally, let λ denote the iteration index of the two-step algorithm. The algorithm stops when one of the following conditions is satisfied: the parameter estimates stabilize, i.e., $\mathbf{\Theta}^{(\lambda)} = \mathbf{\Theta}^{(\lambda-1)}$; or the improvement in the log-likelihood function falls below a predefined threshold, i.e., $\left|\mathcal{L}^{(\lambda)} - \mathcal{L}^{(\lambda-1)}\right| \leq \gamma$, where $\gamma > 0$; or the maximum number of iterations $\lambda_{\max}$ is reached. These stopping rules ensure convergence of the procedure while avoiding unnecessary computational burden.

5.2 QUASI MAXIMUM LIKELIHOOD ESTIMATION

To estimate the parameters of our dynamic spatial panel model with group-specific heterogeneity, we adopt a Quasi-Maximum Likelihood approach. This methodology builds on the framework proposed by Yu et al. (2008); Lee and Yu (2010a,b) for spatial dynamic panel model, when cluster heterogeneity is not considered, and implemented by Simonovska (2025). The model is split into two conceptual blocks: one capturing the standard (non-spatial) dependence of the dependent variable on covariates, and another focusing on the spatially lagged dependent variable and the effects of covariates on it.

In their approach, spatial dependence is treated as a correction to the primary regression, allowing for sequential estimation: first by estimating the effect of covariates without spatial feedback, and then adjusting for spatial spillovers using a residual correction term.

Our method extends this idea to the more complex setting of heterogeneous spatial dependence. Specifically, instead of a single spatial component, we consider multiple group-to-group spatial dependencies, with one spatial autoregressive coefficient for each pair of groups. We decompose the model into a non-spatial block and several spatial blocks, each corresponding to a specific group interaction. This block-wise structure allows us to first estimate the base regression, then separately estimates the effects of regressors on the spatial lag of the dependent variable for each block, and finally recover the spatial autoregressive coefficients by maximizing a concentrated log-likelihood function that accounts for the entire spatial dependence structure.

This modular procedure not only simplifies estimation in the presence of heterogeneity but also improves computational tractability and interpretability of the estimated parameters.

Let $Y_t = (y_{1t}, \dots, y_{Nt})'$ denote the $N \times 1$ vector of dependent variables at time t, and $X_t = (\mathbf{x}'_{1t}, \dots, \mathbf{x}'_{Nt})'$ the $N \times k$ matrix of covariates. Let $W \in \mathbb{R}^{N\times N}$ be the spatial weights matrix and $P \in \{0,1\}^{G\times N}$ the group membership matrix, where each column of P contains a single entry equal to 1, indicating the group to which each unit belongs.

Regarding the spatial weights matrix W, in this paper we assume a single spatial structure that governs all spatial and spatiotemporal spillover effects. However, the framework is flexible and can accommodate multiple spatial weight matrices, such as W_x and W_y, to separately capture spatial spillovers in the independent and dependent variables, or matrices like W_0 and W_1, which distinguish between contemporaneous and temporally lagged spillovers.

We aim to estimate the set of regression parameters

$$\boldsymbol{\Theta} = (\boldsymbol{\tau}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\rho}, \sigma)$$

From the model described in Equation 5, which can be expressed as:

$$\begin{aligned}\left(I_N - P^\top \boldsymbol{\rho} P \circ W\right) Y_t = \operatorname{diag}(P^\top \boldsymbol{\tau}) Y_{t-1} + \left(P^\top \boldsymbol{\eta} P \circ W\right) Y_{t-1} \\ + \operatorname{diag}_k(P^\top \boldsymbol{\beta}) X_t + \left(P^\top \boldsymbol{\delta} P \circ W\right) X_t + \epsilon_t,\end{aligned} \tag{7}$$

where the Hadamard product "$\circ$" captures group-specific spatial interactions. We recall that we are not considering any fixed effects and that the error variance is constant over the sample. To simplify notation, we collect all regressors and lagged terms into a single matrix Z_t and rewrite the model as:

$$\left(I_N - P^\top \boldsymbol{\rho} P \circ W\right) Y_t = Z_t \mathbf{B} + \epsilon_t, \tag{8}$$

where Z_t is the $N \times q$ matrix of explanatory variables which includes the temporal lag of the dependent variable, the lagged spatial spillover of the dependent variable, the covariates and their spillovers, all interacted by the clustering membership. Thus defined as

$$Z_t = \left[\, Y_{t-1},\ (P^\top \boldsymbol{\eta} P \circ W) Y_{t-1},\ X_t,\ (P^\top \boldsymbol{\delta} P \circ W) X_t \,\right],$$

and $\mathbf{B}$ is the stacked parameter vector, with lenght q, including all the regression coefficient of the model in Equation 5, except $\boldsymbol{\rho}$, thus defined as

$$\mathbf{B}^\top = \left[\, \boldsymbol{\tau}^\top,\ \operatorname{vec}(\boldsymbol{\eta})^\top,\ \boldsymbol{\beta}^\top,\ \operatorname{vec}(\boldsymbol{\delta})^\top \,\right].$$

Focusing on the left-hand side (LHS), we can equivalently express the spatial dependence as:

$$\left(I_N - \sum_{g=1}^{G} \sum_{f=1}^{G} \rho_{gf} W^{(gf)} \right) Y_t = Z_t \mathbf{B} + \epsilon_t, \tag{9}$$

where $W^{(gf)} \in \mathbb{R}^{N \times N}$ denotes a group-structured spatial matrix, defined elementwise as

$$w_{ij}^{(gf)} = \begin{cases} w_{ij}, & \text{if } i \in \mathcal{C}_g \text{ and } j \in \mathcal{C}_f, \\ 0, & \text{otherwise.} \end{cases}$$

Here, $\mathcal{C}_g$ denotes the set of units belonging to group g. Each matrix $W^{(gf)}$ captures spatial connections only between units in groups g and f, while maintaining the full $N \times N$ dimension.

Remark. Notice that $W^{(gf)}$ do not correspond to the submatrix W_{gf}. The latter would represent the block of W corresponding to rows in $\mathcal{C}_g$ and columns in $\mathcal{C}_f$, hence its dimensions depend on the size of clusters g and f. In contrast, $W^{(gf)}$ preserves the full $N \times N$ structure, with nonzero entries only where both group indices match, which ensures algebraic compatibility in Hadamard products and summations.

We now partition the model into its non-spatial and spatial components:

$$Y_t = Z_t \mathbf{B}_{ns} + \epsilon_{t,ns}, \tag{10}$$

$$-\rho_{gf} W^{(gf)} Y_t = Z_t \mathbf{B}_{sp}^{(gf)} + \epsilon_{t,sp}^{(gf)} \qquad \forall g, f = 1, \ldots, G \tag{11}$$

where $\mathbf{B}_{ns}$ collects the non-spatial parameters and $\mathbf{B}_{sp}^{(gf)}$ captures the contribution of the (g, f)-specific spatial effect, thus

$$\mathbf{B}_{ns}^\top = \left[\, \boldsymbol{\tau}_{ns}^\top,\ \mathrm{vec}(\boldsymbol{\eta}_{ns})^\top,\ \boldsymbol{\beta}_{ns}^\top,\ \mathrm{vec}(\boldsymbol{\delta}_{ns})^\top \,\right],$$
$$\mathbf{B}_{sp}^{(gf)}\top = \left[\, \boldsymbol{\tau}_{sp}^{(gf)}\top,\ \mathrm{vec}(\boldsymbol{\eta})_{sp}^{(gf)}\top,\ \boldsymbol{\beta}_{sp}^{(gf)}\top,\ \mathrm{vec}(\boldsymbol{\delta})_{sp}^{(gf)}\top \,\right], \quad \forall g, f = 1, \ldots, G.$$

Remark. Notice that, the sum of the RHS of Equations 10 and 11, correspond to the RHS of Equation 9, thus sum of the LHS of Equations 10 and 11, correspond to the LHS of Equation 9. By construction,

$$\mathbf{B} = \mathbf{B}_{ns} + \sum_{g=1}^{G} \sum_{f=1}^{G} \mathbf{B}_{sp}^{(gf)}, \qquad \epsilon_t = \epsilon_{t,ns} + \sum_{g=1}^{G} \sum_{f=1}^{G} \epsilon_{t,sp}^{(gf)}, \quad \forall g, f = 1, \ldots, G.$$

To facilitate estimation of the spatial blocks, we define:

$$W^{(gf)} Y_t = Z_t \mathbf{B}_{sp}^{*(gf)} + \epsilon_{t,sp}^{*(gf)}, \quad \forall g, f = 1, \ldots, G$$

such that

$$\mathbf{B}_{sp}^{(gf)} = -\rho_{gf} \mathbf{B}_{sp}^{*(gf)}, \qquad \epsilon_{t,sp}^{(gf)} = -\rho_{gf} \epsilon_{t,sp}^{*(gf)}, \quad \forall g, f = 1, \ldots, G.$$

The vectors of coefficients in the regression blocks $\mathbf{B}_{ns}$ and $\mathbf{B}_{sp}^{*(gf)}$ can then be estimated using least squares as:

$$\hat{\mathbf{B}}_{ns} = (Z'Z)^{-1} Z'Y,$$
$$\hat{\mathbf{B}}_{sp}^{*(gf)} = (Z'Z)^{-1} Z' W_T^{(gf)} Y, \quad \forall g, f = 1, \ldots, G$$

where Y is the $NT \times 1$ stacked vector of outcomes, Z is the $NT \times q$ regressor matrix, and $W_T = I_T \otimes W$, $W_T^{(gf)} = I_T \otimes W^{(gf)}$.

The corresponding residuals are:

$$\hat{\epsilon}_{ns} = Y - Z\hat{\mathbf{B}}_{ns}, \qquad \hat{\epsilon}_{sp}^{*(gf)} = W_T^{(gf)}Y - Z\hat{\mathbf{B}}_{sp}^{*(gf)}.$$

We then maximize the quasi log-likelihood function from Yu et al. (2008); Lee and Yu (2010a) with respect to $\boldsymbol{\rho}$, given by:

$$\mathcal{L}(Y|Z, W, \boldsymbol{\rho}, \mathcal{P}^G) \propto -\frac{NT_0}{2}\log\left(\frac{1}{NT_0}\sum_{t=1}^{T_0}\epsilon_t'\epsilon_t\right) + T_0\log\left|I_N - \sum_{g=1}^{G}\sum_{f=1}^{G}\rho_{gf}W^{(gf)}\right|, \tag{12}$$

where $T_0 = T$ if the model is not dynamic, and $T_0 = T - 1$ if the model is dynamic. The total residual ϵ_t can be found as the sum of the residual from the non-spatial regression block and the residuals from the spatial regression blocks, thus

$$\epsilon_t = \hat{\epsilon}_{t,ns} + \sum_{g=1}^{G}\sum_{f=1}^{G}\epsilon_{t,sp}^{(gf)} = \hat{\epsilon}_{t,ns} - \sum_{g=1}^{G}\sum_{f=1}^{G}\rho_{gf}\hat{\epsilon}_{t,sp}^{*(gf)}$$

and thus, the squared of the total residuals can be obtained as the square of the binomial as follows

$$\begin{aligned}\epsilon_t'\epsilon_t &= \left(\hat{\epsilon}_{t,ns} - \sum_{g=1}^{G}\sum_{f=1}^{G}\rho_{gf}\hat{\epsilon}_{t,sp}^{*(gf)}\right)'\left(\hat{\epsilon}_{t,ns} - \sum_{g=1}^{G}\sum_{f=1}^{G}\rho_{gf}\hat{\epsilon}_{t,sp}^{*(gf)}\right)\\ &= \hat{\epsilon}_{t,ns}'\hat{\epsilon}_{t,ns} - 2\hat{\epsilon}_{t,ns}'\sum_{g=1}^{G}\sum_{f=1}^{G}\rho_{gf}\hat{\epsilon}_{t,sp}^{(gf)} + \left(\sum_{g=1}^{G}\sum_{f=1}^{G}\rho_{gf}\hat{\epsilon}_{t,sp}^{(gf)}\right)'\left(\sum_{g=1}^{G}\sum_{f=1}^{G}\rho_{gf}\hat{\epsilon}_{t,sp}^{(gf)}\right).\end{aligned} \tag{13}$$

Plugging the residuals in 13 in the likelihood function in 12, it is possible to estimated spatial parameters $\boldsymbol{\rho}$ as

$$\hat{\boldsymbol{\rho}} = \arg\max_{\boldsymbol{\rho}}\mathcal{L}(Y|Z, W, \boldsymbol{\rho}, \mathcal{P}^G).$$

Given $\hat{\boldsymbol{\rho}}$, we update the coefficients and residuals as:

$$\hat{\mathbf{B}} = \hat{\mathbf{B}}_{ns} - \sum_{g=1}^{G}\sum_{f=1}^{G}\hat{\rho}_{gf}\,\hat{\mathbf{B}}_{sp}^{*(gf)},$$

$$\hat{\epsilon}_t = \hat{\epsilon}_{t,ns} - \sum_{g=1}^{G}\sum_{f=1}^{G} \hat{\rho}_{gf}\, \hat{\epsilon}^{*(gf)}_{t,sp}.$$

Finally, the fitted values for the entire panel are given by:

$$\widehat{Y} = \left(I_{NT} - \sum_{g=1}^{G}\sum_{f=1}^{G} \hat{\rho}_{gf} W_T^{(gf)} \right)^{-1} Z\widehat{\mathbf{B}},$$

and the error variance is computed as:

$$\widehat{\sigma}^2_{\varepsilon} = \frac{\widehat{\epsilon}'\widehat{\epsilon}}{NT}.$$

To conduct inference on the estimated parameters, we compute their asymptotic variance–covariance matrix under the quasi–maximum likelihood (QML) framework. Following Yu et al. (2008); Lee and Yu (2010a), the variance of $\hat{\mathbf{\Theta}}$ can be written in "sandwich" form as

$$\widehat{\mathrm{Var}}(\hat{\mathbf{\Theta}}) = \widehat{\Sigma}^{-1} + \widehat{\Sigma}^{-1}\widehat{\Omega}\,\widehat{\Sigma}^{-1},$$

where $\widehat{\Sigma}$ is the (expected) information matrix evaluated at $(\hat{\mathbf{B}}, \hat{\boldsymbol{\rho}}, \hat{\sigma}^2)$, and $\widehat{\Omega}$ is a correction term that accounts for possible departures from Gaussianity (excess kurtosis) in the innovations. In the multi-parameter spatial setting, let

$$S(\boldsymbol{\rho}) = I_N - \sum_{g=1}^{G}\sum_{f=1}^{G} \rho_{gf} W^{(gf)}, \qquad G_{gf}(\boldsymbol{\rho}) = W^{(gf)} S(\boldsymbol{\rho})^{-1},$$

so that $\widehat{\Sigma}$ is assembled using $S(\hat{\boldsymbol{\rho}})^{-1}$ and the derivative matrices $G_{gf}(\hat{\boldsymbol{\rho}})$ that arise from differentiation of $\log|S(\boldsymbol{\rho})|$ and of the quadratic form in the QML objective. Let $\widehat{V}(\hat{\mathbf{\Theta}})$ denote the resulting variance–covariance matrix; standard errors are then obtained as the square roots of its diagonal entries.

Given $\widehat{V}(\hat{\mathbf{\Theta}})$, inference for each component Θ_j of $\mathbf{\Theta}$ is based on the asymptotic normal approximation. For each parameter estimate $\hat{\Theta}_j$, we compute its standard error

$$se(\hat{\Theta}_j) = \sqrt{\widehat{V}_{jj}},$$

form the corresponding Wald statistic

$$z_j = \frac{\hat{\Theta}_j}{se(\hat{\Theta}_j)},$$

and obtain a two-sided p-value as

$$p_j = 2\Phi(-|z_j|)\,,$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. This procedure is applied to all elements of $\hat{\mathbf{B}}$ as well as to each group-to-group spatial coefficient $\hat{\rho}_{gf}$ (and, when of interest, to $\hat{\sigma}^2$).

6 SIMULATION EXPERIMENT

To evaluate the proposed estimation procedure, we design a Monte Carlo experiment based on the group-structured dynamic spatial panel data model introduced in Section 2. The general specification in Equation (2) includes several layers of spatial and temporal dependence. However, as discussed in Anselin et al. (2008) and Elhorst (2010), the full model suffers from identification issues when all dependence parameters are simultaneously included. For this reason, applied research commonly relies on restricted versions that incorporate specific combinations of parameters.

Following this approach, we design a series of data-generating processes (DGPs) that consider different subsets of the full parameter space, enabling us to separately assess the performance of the estimator under various spatial and dynamic configurations. In particular, we consider five restricted models (M1–M5), each including at least one of the structural parameters of interest—spatial, temporal, or covariate effects—but never all simultaneously.

6.1 SPATIAL STRUCTURE AND HETEROGENEITY SETTINGS

The spatial matrix W is constructed using a k-nearest neighbors approach with $k = 10$, based on random coordinates uniformly distributed in $[0, 1]^2$. This matrix is row-standardized and has zeros on the diagonal.

We introduce heterogeneity in the spatial dependence through different configurations of the group-to-group dependence matrix $\boldsymbol{\psi} = (\boldsymbol{\rho}, \boldsymbol{\eta}, \boldsymbol{\delta})$, representing three distinct levels of cross-cluster interaction (see Section 3):

- H1 - No Cross-Group Spillovers: heterogeneous within clusters but no spatial link across clusters.

$$\boldsymbol{\psi} = [\psi_{gf}] = \begin{bmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{bmatrix} \quad \text{H1}$$

- H2 - Homogeneous Cross-Group Spillovers: heterogeneous within clusters and symmetric across clusters (i.e., $\rho_{12} = \rho_{21}$).
$$\boldsymbol{\psi} = [\psi_{gf}] = \begin{bmatrix} \psi_1 & \psi_{12} \\ \psi_{12} & \psi_2 \end{bmatrix} \quad \text{H2}$$
- H3 - Asymmetric Cross-Group Spillovers: heterogeneous within clusters and asymmetric across clusters ($\rho_{12} \neq \rho_{21}$).
$$\boldsymbol{\psi} = [\psi_{gf}] = \begin{bmatrix} \psi_1 & \psi_{12} \\ \psi_{21} & \psi_2 \end{bmatrix} \quad \text{H3}$$

In all settings, within-cluster dependencies (ρ_{gg}) differ across groups. The true parameter values used in the simulations are:

$$\boldsymbol{\rho} = \begin{bmatrix} \rho_1 & \rho_{12} \\ \rho_{21} & \rho_2 \end{bmatrix} \begin{cases} \begin{bmatrix} 0.25 & 0 \\ 0 & 0.15 \end{bmatrix} \\ \begin{bmatrix} 0.25 & 0.05 \\ 0.05 & 0.15 \end{bmatrix} \\ \begin{bmatrix} 0.25 & 0.05 \\ 0.02 & 0.15 \end{bmatrix}. \end{cases}$$

Other parameters are set as follows:

$$\boldsymbol{\tau} = (\tau_1, \tau_2) = (0.3, 0.2), \quad \boldsymbol{\beta} = (\beta_1, \beta_2) = (0.4, 0.6), \quad \boldsymbol{\eta} = 0.4\boldsymbol{\rho}, \quad \boldsymbol{\delta} = 0.6\boldsymbol{\rho}.$$

6.2 MODEL VARIANTS

We simulate data under five model configurations, each corresponding to a different restriction of the general framework:

- **M1**: Spatial lag ρ and covariate β
$$Y_t = \left(P^\top \boldsymbol{\rho} P \circ W\right) Y_t + \text{diag}_k(P^\top \boldsymbol{\beta}) X_t + \epsilon_t,$$
- **M2**: Spatial lag ρ and temporal lag τ
$$Y_t = \left(P^\top \boldsymbol{\rho} P \circ W\right) Y_t + \text{diag}(P^\top \boldsymbol{\tau}) Y_{t-1} + \epsilon_t,$$
- **M3**: Spatial Durbin Model with ρ, β, and spatial lag of covariates δ
$$Y_t = \left(P^\top \boldsymbol{\rho} P \circ W\right) Y_t + \text{diag}_k(P^\top \boldsymbol{\beta}) X_t + \left(P^\top \boldsymbol{\delta} P \circ W\right) X_t + \epsilon_t,$$
- **M4**: Spatiotemporal autoregressive model with ρ, τ, and lagged spatial outcome η
$$Y_t = \left(P^\top \boldsymbol{\rho} P \circ W\right) Y_t + \text{diag}(P^\top \boldsymbol{\tau}) Y_{t-1} + \left(P^\top \boldsymbol{\eta} P \circ W\right) Y_{t-1} + \epsilon_t,$$

- **M5**: Spatiotemporal autoregressive model with covariate, with ρ, τ, η, β

$$Y_t = \left(P^\top \boldsymbol{\rho} P \circ W\right) Y_t + \mathrm{diag}(P^\top \boldsymbol{\tau}) Y_{t-1} + \left(P^\top \boldsymbol{\eta} P \circ W\right) Y_{t-1} + \mathrm{diag}_k(P^\top \boldsymbol{\beta}) X_t + \epsilon_t,$$

Each specification is therefore a restricted but identifiable form of the general model, allowing us to evaluate how estimation accuracy changes with model complexity.

6.3 MONTE CARLO IMPLEMENTATION

Each configuration is repeated over $MC = 100$ Monte Carlo iterations with $N = 200$ spatial units, $T = 50$ time periods, and $G = 2$ clusters. The relative size of the two clusters varies across simulations: for each iteration, a probability 0.5 is drawn from a uniform distribution $U(0.2, 0.8)$, and each unit i is assigned to cluster 1 with probability p_g and to cluster 2 with probability $1 - p_g$. This ensures that cluster sizes are random but balanced on average.

Simulation outputs include estimated spatial parameters (ρ_{gf}), temporal and dynamic lags (τ_g, η_{gf}), covariate effects (β_g, δ_{gf}), residual variance σ^2_ϵ. Moreover, for models that include the covariate variable X we are able to estimates direct and indirect effects over the entire sample, as well as cluster-specific direct effects and cross-clusters indirect effects.

Algorithm 1 Monte Carlo Simulation Procedure

Set $N = 100$, $T = 50$, $G = 2$, $k = 1$, $MC = 100$
Simulate a clustering partition, $\mathcal{P}^{G=2}$
Construct W via k-nearest neighbors ($k = 10$)
Set parameters $\boldsymbol{\Theta}$
Compute true values of direct and indirect effects within clusters and across clusters
for $mc = 1$ to MC **do**
  for each heterogeneity configuration (H1–H3) **do**
    Condiser the proper subset of parameters according to the heterogeneity configuration
    for each model specification (M1–M5) **do**
      Condiser the proper subset of parameters according to the model specification
      Simulate panel data (Y_t, X_t) using the DGP, such that $Y_t = f(X_t, \Theta, \mathcal{P}^2, W)$
      Estimate model parameters $\widehat{\boldsymbol{\Theta}}$ using QML procedure
      Compute estimates of direct and indirect effects within clusters and across clusters
      Store estimated coefficients, direct and indirect effects
    end for
  end for
end for

This design allows us to systematically evaluate how different sources of spatial heterogeneity and model restrictions affect estimation performance and parameter recovery.

6.4 DISCUSSION OF SIMULATION RESULTS

Overall, the simulation results confirm that the proposed estimation procedure performs well in recovering the heterogeneous spatial and dynamic parameters across different model specifications and heterogeneity settings. Figure 1 show the distribution of the estimates of spatial and temporal dependence parameters considering models M1-M5 respectively in panels a-e, and considering different heterogeneity structures of the parameters H1-H3 as described in previously. In each boxplot, the white circle indicates the true value of the parameters.

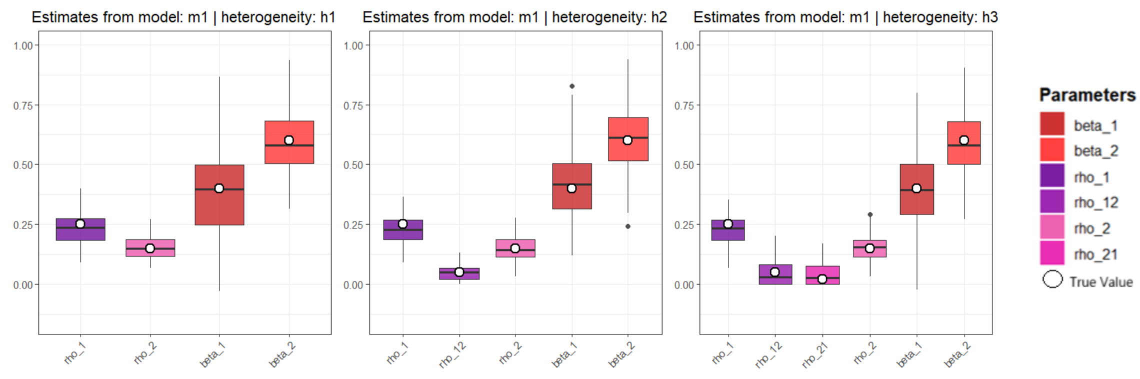
The estimates of the spatial autoregressive parameters $\boldsymbol{\rho}$ –represented in shadow of purple– are generally accurate and display low variability, closely reflecting the underlying cluster heterogeneity imposed in the data-generating process. This indicates that the estimation method is able to effectively capture group-specific spatial dependencies.

For the simpler model specifications (M1 and M2), the estimators exhibit good performance, with the majority of estimated parameters centered around their true values. Although the boxplots of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ display higher variability (i.e., longer boxes), their medians correctly reflect the cluster-specific heterogeneity: the estimated τ_g parameters are higher for cluster 1 and lower for cluster 2, while the opposite pattern is observed for the β_g parameters, which are higher for cluster 2 and lower for cluster 1. This consistency across replications suggests that the method reliably distinguishes between cluster-specific effects even in the presence of moderate dispersion.

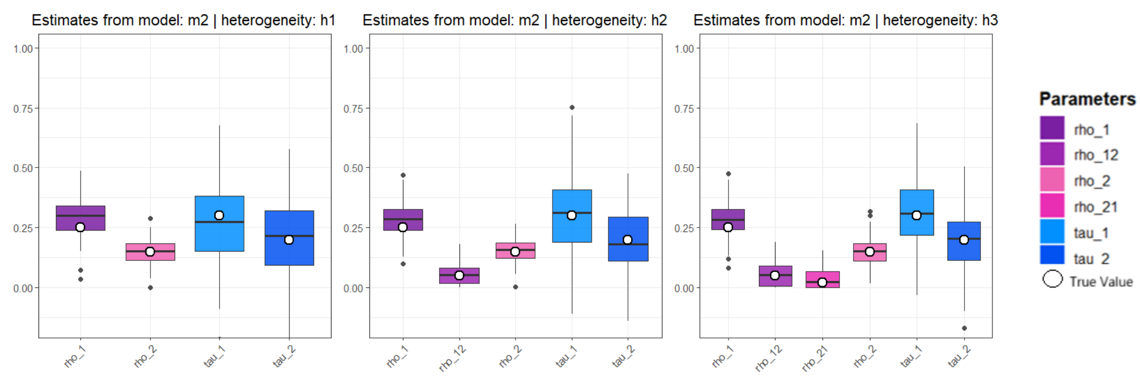
In contrast, for the more complex models (M3–M5), which include multiple spatial effects ($\boldsymbol{\rho}, \boldsymbol{\delta}$ and $\boldsymbol{\eta}$), the estimation procedure shows some limitations. In particular, the spatiotemporal coefficients δ_{gf} and the lagged spatial effects η_{gf} tend to be slightly overestimated, while the autoregressive spatial parameters ρ_{gf} are somewhat underestimated. This bias is likely due to the increased complexity of jointly identifying several interacting spatial components. Nonetheless, the model remains capable of capturing the relative heterogeneity structure across clusters: the estimated parameters consistently follow the ordering $\psi_1 > \psi_2 > \psi_{12} \geq \psi_{21}$, as in the settings, meaning that within-cluster effects in group 1 are generally stronger than those in group 2, while cross-cluster effects are smaller and slightly asymmetric, consistently with the true values of the parameters.

Now, we consider the resulting direct and indirect effects of the independent variable in models M1, M3 and M5. Direct and Indirect effects have been computed as described in Section 4, considering cluster-specifics effects and cross-clusters effects.

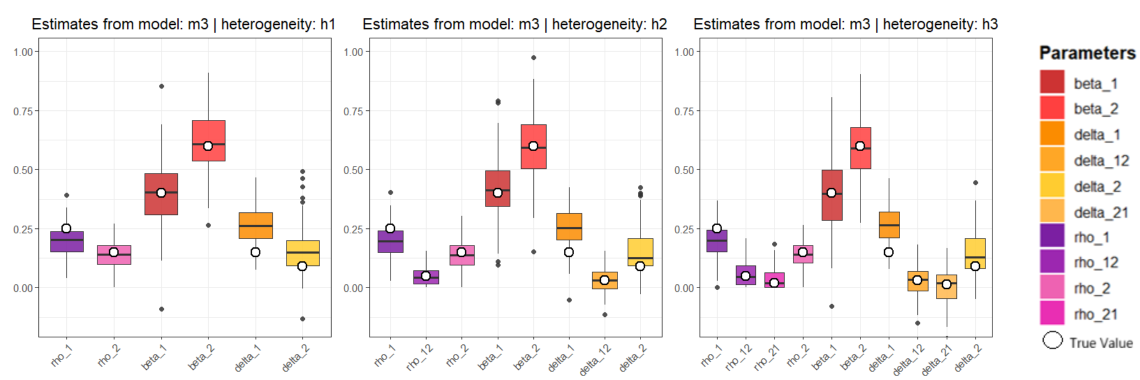
Figure 2 illustrates the distribution of the estimated effects of X on Y across Monte Carlo replications, broken down by model and heterogeneity scenario. Each boxplot



(a) Simulation results – Model 1

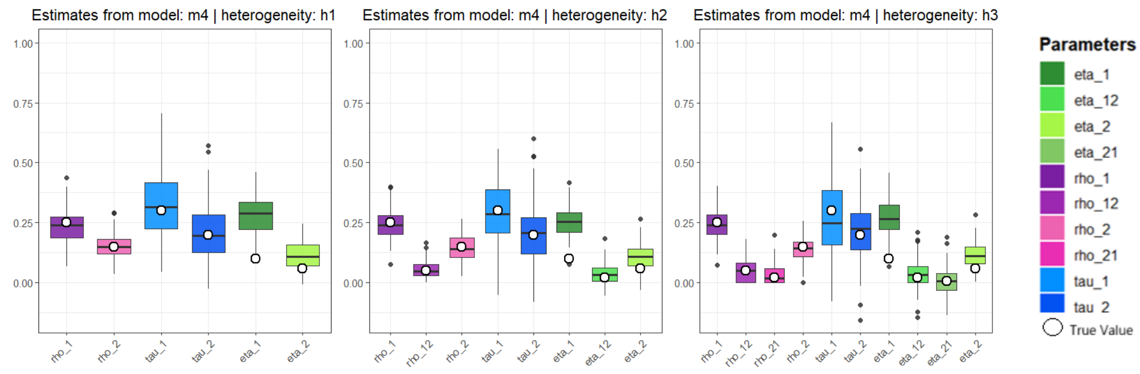


(b) Simulation results – Model 2

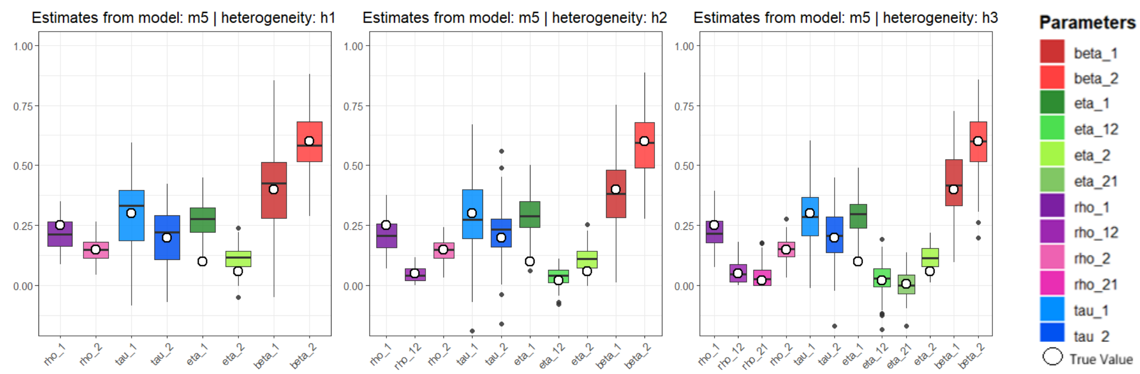


(c) Simulation results – Model 3

Figure 1: Monte Carlo simulation results (a–c) for models M1–M3 under various heterogeneity settings. The boxplots show the distribution of the estimates of spatial and temporal dependence parameters, while the white circle indicates the true value of the parameters. Panels continue on next page.



(d) Simulation results – Model 4

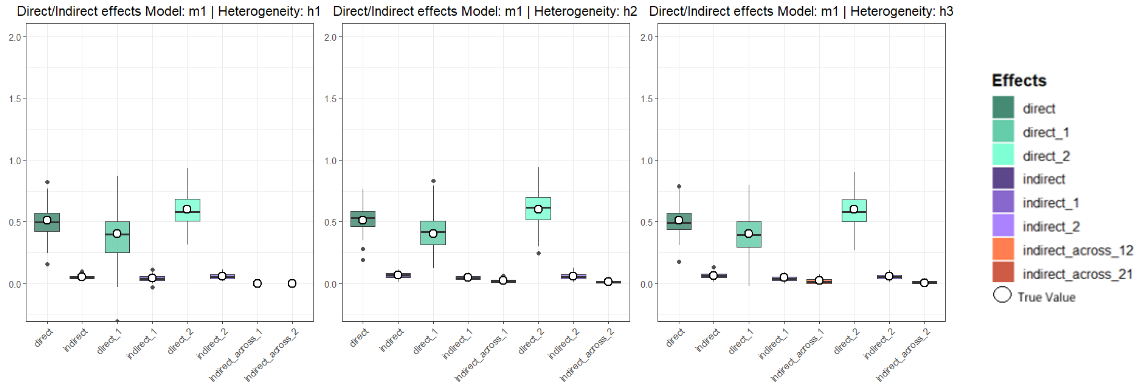


(e) Simulation results – Model 5

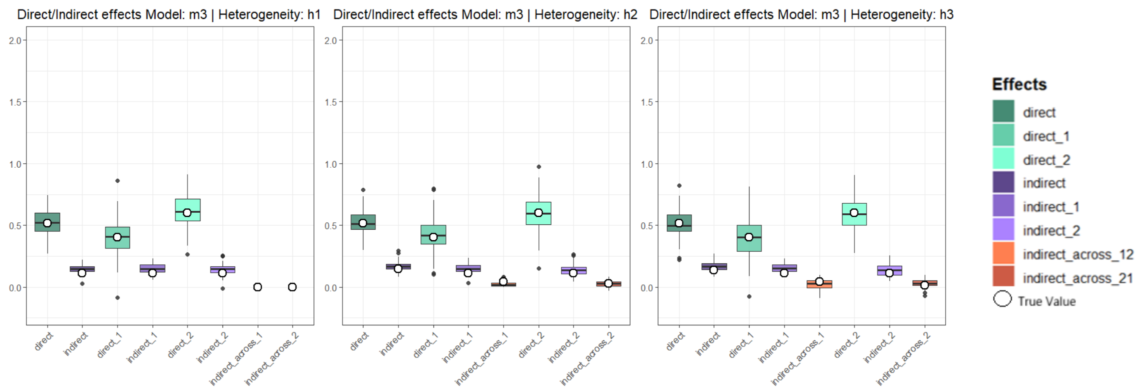
Figure 1: (continued) Monte Carlo simulation results (d–e) for models M4–M5 under various heterogeneity settings. The boxplots show the distribution of the estimates of spatial and temporal dependence parameters, while the white circle indicates the true value of the parameters.

corresponds to a different spatial effect component: total, cluster-specific, and cross-cluster direct and indirect effects. The horizontal axis labels denote the specific type of effect, where "direct" and "indirect" refer to the global effects, "dir_1" and "indir_1" refer to the effects in clusters 1, "ind_across_1" refers to the resulting effect on Y in units in cluster 1, from a change of X in a units in cluster 2, similarly "dir_2" and "indir_2" refer to the effects in clusters 2, "ind_across_2" refers to the resulting effect on Y in units in cluster 2, from a change of X in a units in cluster 1. The true value of each effect is overlaid as a white circle, allowing visual assessment of estimation accuracy and bias.

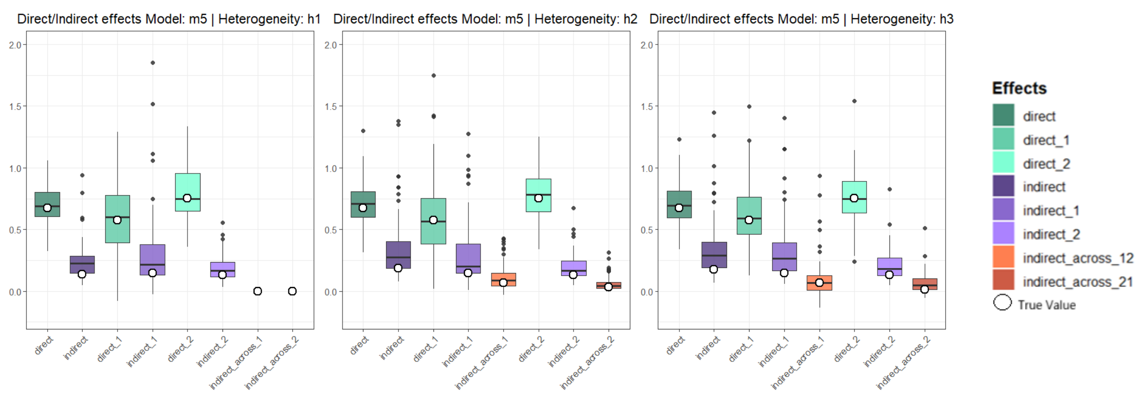
Overall, the estimation procedure demonstrates strong performance. In most cases, the boxplots are centered closely around the true effect values, indicating unbiasedness. For example, in Model M1, the white circles often lie near the center of the interquartile



(a) Estimated effects – Model 1



(b) Estimated effects – Model 3



(c) Estimated effects – Model 5

Figure 2: Boxplots of estimated direct and indirect effects for models M1, M3, and M5 under different heterogeneity structures. The boxplots show the distribution of the estimates of general direct and indirect effects, cluster-specific and cross-cluster direct and indirect effects. The white circle indicates the true value of the effects.

range, especially for the direct and indirect effects of clusters 1 and 2, suggesting accurate estimation and moderate variability.

Notably, the model captures the cluster-specific heterogeneity of the effects well. Across all models and heterogeneity settings, we observe that the estimated direct and indirect effects for cluster 1 tend to be higher than those for cluster 2, consistently reflecting the true pattern. This relationship holds not only for the within-cluster effects but also for the cross-cluster spillovers, and is reflected in the estimated values.

In Models M3 and M5, the variability of some estimates increases, especially for indirect effects, as reflected in taller boxplots. Moreover, in model M5 the indirect effects in both clusters seem to be slightly overestimated. Nonetheless, even in these more complex settings, the estimates tend to preserve the correct ranking across clusters, with larger effects being associated with higher estimates. This is particularly important, as it confirms the model's ability to distinguish between different spatial dynamics across groups, even when absolute precision slightly decreases.

Table 1-5 reports simulation results of estimates of model M1-M5 with different heterogeneity structures H1-H3. For each parameter to be estimated in each model, we report its true value in the DGPs, the Mean of the estimates, the Mean Absolute Error (MAE) and the Mean Squared Error (MSE).

Simulation results for Model M1

		Heter. H1			**Heter. H2**			**Heter. H3**		
	True	**Mean**	**MAE**	**MSE**	**Mean**	**MAE**	**MSE**	**Mean**	**MAE**	**MSE**
ρ_1	0.25	0.231	0.055	0.004	0.233	0.050	0.004	0.224	0.051	0.004
ρ_{12}	0.05	–	–	–	0.048	0.027	0.001	0.048	0.041	0.002
ρ_{21}	0.02	–	–	–	–	–	–	0.041	0.036	0.002
ρ_2	0.15	0.151	0.038	0.002	0.149	0.041	0.002	0.151	0.040	0.002
β_1	0.40	0.374	0.137	0.032	0.407	0.113	0.020	0.395	0.122	0.025
β_2	0.60	0.593	0.102	0.015	0.609	0.110	0.018	0.588	0.111	0.018

Table 1: Simulation results of estimates of model M1 with different heterogeneity structures H1-H3. For each parameter to be estimated in the model, we report its true value in the DGPs, the Mean of the estimates, the Mean Absolute Error (MAE) and the Mean Squared Error (MSE).

Simulation results for Model M2

	True	Heter. H1 Mean	MAE	MSE	Heter. H2 Mean	MAE	MSE	Heter. H3 Mean	MAE	MSE
ρ_1	0.25	0.291	0.070	0.008	0.285	0.065	0.006	0.280	0.067	0.007
ρ_{12}	0.05	–	–	–	0.056	0.037	0.002	0.057	0.043	0.003
ρ_{21}	0.02	–	–	–	–	–	–	0.040	0.037	0.002
ρ_2	0.15	0.149	0.043	0.003	0.156	0.039	0.002	0.152	0.046	0.003
τ_1	0.30	0.275	0.129	0.027	0.314	0.129	0.028	0.306	0.120	0.023
τ_2	0.20	0.211	0.128	0.025	0.188	0.110	0.019	0.195	0.097	0.016

Table 2: Simulation results of estimates of model M2 under different heterogeneity structures H1–H3. For each parameter, we report its true value used in the DGP, the Mean of the estimates, the Mean Absolute Error (MAE), and the Mean Squared Error (MSE).

Simulation results for Model M3

	True	Heter. H1 Mean	MAE	MSE	Heter. H2 Mean	MAE	MSE	Heter. H3 Mean	MAE	MSE
ρ_1	0.25	0.196	0.070	0.008	0.201	0.071	0.007	0.198	0.058	0.006
ρ_{12}	0.05	–	–	–	0.047	0.030	0.001	0.058	0.043	0.002
ρ_{21}	0.02	–	–	–	–	–	–	0.039	0.037	0.002
ρ_2	0.15	0.137	0.051	0.004	0.136	0.049	0.004	0.137	0.046	0.004
β_1	0.40	0.399	0.113	0.021	0.416	0.106	0.019	0.397	0.123	0.025
β_2	0.60	0.615	0.100	0.016	0.594	0.114	0.021	0.593	0.104	0.016
δ_1	0.15	0.259	0.116	0.019	0.253	0.114	0.018	0.269	0.122	0.019
δ_{12}	0.03	–	–	–	0.030	0.041	0.003	0.028	0.051	0.004
δ_{21}	0.012	–	–	–	–	–	–	0.005	0.057	0.005
δ_2	0.09	0.154	0.084	0.014	0.148	0.083	0.013	0.144	0.078	0.011

Table 3: Simulation results of estimates of model M3 under different heterogeneity structures H1–H3. For each parameter, we report its true value used in the DGP, the Mean of the estimates, the Mean Absolute Error (MAE), and the Mean Squared Error (MSE).

Simulation results for Model M4

	True	Heter. H1 Mean	MAE	MSE	Heter. H2 Mean	MAE	MSE	Heter. H3 Mean	MAE	MSE
ρ_1	0.25	0.238	0.059	0.006	0.243	0.052	0.004	0.240	0.049	0.004
ρ_{12}	0.05	–	–	–	0.054	0.030	0.001	0.051	0.032	0.001
ρ_{21}	0.02	–	–	–	–	–	–	0.034	0.031	0.002
ρ_2	0.15	0.153	0.039	0.003	0.145	0.043	0.003	0.140	0.038	0.002
τ_1	0.30	0.324	0.113	0.020	0.297	0.108	0.018	0.268	0.131	0.024
τ_2	0.20	0.213	0.098	0.016	0.212	0.107	0.019	0.220	0.103	0.017
η_1	0.10	0.282	0.182	0.040	0.257	0.157	0.028	0.274	0.174	0.036
η_{12}	0.02	–	–	–	0.037	0.037	0.002	0.032	0.046	0.004
η_{21}	0.008	–	–	–	–	–	–	0.005	0.048	0.004
η_2	0.06	0.115	0.063	0.006	0.108	0.061	0.006	0.111	0.061	0.006

Table 4: Simulation results of estimates of model M4 under different heterogeneity structures H1–H3. For each parameter, we report its true value used in the DGP, the Mean of the estimates, the Mean Absolute Error (MAE), and the Mean Squared Error (MSE).

Simulation results for Model M5

	True	Heter. H1 Mean	MAE	MSE	Heter. H2 Mean	MAE	MSE	Heter. H3 Mean	MAE	MSE
ρ_1	0.25	0.214	0.061	0.005	0.208	0.066	0.006	0.221	0.057	0.005
ρ_{12}	0.05	–	–	–	0.046	0.027	0.001	0.059	0.040	0.002
ρ_{21}	0.02	–	–	–	–	–	–	0.042	0.037	0.002
ρ_2	0.15	0.149	0.040	0.003	0.146	0.041	0.002	0.153	0.039	0.002
τ_1	0.30	0.292	0.120	0.014	0.294	0.128	0.025	0.293	0.102	0.016
τ_2	0.20	0.205	0.102	0.014	0.221	0.084	0.012	0.206	0.099	0.015
η_1	0.10	0.274	0.175	0.035	0.291	0.192	0.041	0.293	0.193	0.043
η_{12}	0.02	–	–	–	0.037	0.036	0.002	0.027	0.050	0.004
η_{21}	0.008	–	–	–	–	–	–	0.005	0.046	0.004
η_2	0.06	0.116	0.063	0.006	0.109	0.066	0.005	0.117	0.065	0.006
β_1	0.40	0.399	0.113	0.021	0.397	0.126	0.028	0.423	0.117	0.021
β_2	0.60	0.607	0.140	0.017	0.595	0.107	0.017	0.590	0.106	0.018

Table 5: Simulation results of estimates of model M5 under different heterogeneity structures H1–H3. For each parameter, we report its true value used in the DGP, the Mean of the estimates, the Mean Absolute Error (MAE), and the Mean Squared Error (MSE).

7 CONCLUSION

This paper develops a novel spatiotemporal clustering framework for panel data that jointly accounts for dynamic dependence, spatial spillovers, and structural heterogeneity, both within and across groups of observational units. Unlike traditional models that impose homogeneity across space, our approach explicitly incorporates endogenous group formation and group-specific spatial dynamics, providing a more realistic and flexible representation of complex interdependencies and diffusion processes over space and time. This methodological advancement is particularly important in empirical contexts where assuming uniform effects across units may lead to misleading conclusions.

Our key contribution lies in the development of a unified estimation framework that identifies both the latent cluster structure and the associated heterogeneous parameters. Specifically, we propose a two-step estimation strategy, where the first stage employs a quasi maximum likelihood (QML) approach to estimate cluster-specific spatial and temporal dependence, along with heterogeneous covariate effects and the second detects group membership. This represents a substantial generalization of existing spatial econometric techniques, allowing for richer forms of heterogeneity—including asymmetric spillovers and cross-cluster interactions—while remaining computationally tractable.

To validate the robustness and effectiveness of the proposed methodology, an extensive Monte Carlo simulation study is carried out. The simulation experiment is designed to assess the model's ability to accurately recover the true spatiotemporal parameters and to disentangle the direct and indirect effects under a variety of heterogeneity scenarios. The results confirm that our method performs well even under complex settings, including highly asymmetric and group-specific interaction structures.

Beyond its methodological contributions, the framework has important practical implications. It provides researchers and practitioners with a powerful tool for uncovering the true underlying structure of complex phenomena, particularly when interactions and dependencies are not uniform across observational units. By identifying how different variables influence one another within and across heterogeneous groups, the model can yield deeper insights into context-specific dynamics. Moreover, the ability to detect and characterize group-specific spatial behavior has significant value for policymakers, who can use such information to design more effective and targeted interventions. For example, in environmental or regional policy, identifying clusters of regions with similar emission patterns or responsiveness to environmental drivers allows for the formulation of tailored, group-specific strategies that are more likely to be effective than one-size-fits-all approaches.

We are currently applying this model to the study of greenhouse gas (GHG) emissions across European regions. The empirical setting leverages a rich dataset of regional emissions, sourced from EDGAR v8.0 and harmonized at the NUTS-2 level through the European Commission's ARDECO database. The data cover 234 regions across 27 EU countries over the period from 1990 to 2022, enabling us to capture both long-run temporal dynamics and fine-grained spatial variation in emissions behavior. Our outcome of interest is total per capita GHG emissions, aggregated over all gases and sectors, with supplementary spatial descriptors such as emissions per km^2 also available.

The spatial structure of the model is informed by a geodetic distance-based spatial weights matrix, allowing us to quantify the role of geographical proximity in shaping regional spillovers. In parallel, the inclusion of socio-economic covariates—such as income, education, sectoral composition, and innovation capacity—permits the exploration of heterogeneous drivers of emissions across different types of regions.

Our ongoing empirical analysis aims to uncover how these socio-economic variables interact with spatial linkages to explain both intra-cluster and inter-cluster spillovers in emissions, and how the marginal effects of key variables differ across structurally distinct groups of regions. Ultimately, this approach offers a more accurate and policy-relevant understanding of regional emission dynamics, which is essential for designing targeted climate policies that account for spatial externalities and structural heterogeneity across regions.

REFERENCES

Luc Anselin and Daniel Arribas-Bel. Spatial fixed effects and spatial dependence in a single cross-section. *Papers in Regional Science*, 92(1):3 – 17, 2013. doi: 10.1111/j.1435-5957.2012.00480.x.

Luc Anselin, Julie Le Gallo, and Hubert Jayet. *Spatial Panel Econometrics*, pages 625–660. 01 2008. ISBN 978-3-540-75889-1. doi: 10.1007/978-3-540-75892-1_19.

Anna Gloria Billé, Roberto Benedetti, and Paolo Postiglione. A two-step approach to account for unobserved spatial heterogeneity. *Spatial Economic Analysis*, 12(4): 452 – 471, 2017. doi: 10.1080/17421772.2017.1286373. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85015661949&doi=10.1080%2f17421772.2017.1286373&partnerID=40&md5=1133d32d83eab33e5903cc0a86dd08d5.

Simone Boccaletti, Paolo Maranzano, and Miguel Viegas. Inequality and concentration in farmland production and size: Regional analysis for the european union from 2010 to 2020, 2024. URL https://arxiv.org/abs/2409.00111.

Stéphane Bonhomme and Elena Manresa. Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147 – 1184, 2015. doi: 10.3982/ECTA11319. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-84930701347&doi=10.3982%2fECTA11319&partnerID=40&md5=85d9497a93302077edaf408c61409777.

Roy Cerqueti, Pierpaolo D'Urso, and Raffaele Mattera. Fuzzy group fixed-effects estimation with spatial clustering. *AStA Advances in Statistical Analysis*, 2025. doi: 10.1007/s10182-024-00522-1. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85214015944&doi=10.1007%2fs10182-024-00522-1&partnerID=40&md5=8c8a11a3d6941ed67685b09b8b096a25.

Jia Chen, Yongcheol Shin, and Chaowen Zheng. Estimation and inference in heterogeneous spatial panels with a multifactor error structure. *Journal of Econometrics*, 229(1):55 – 79, 2022. doi: 10.1016/j.jeconom.2021.05.003. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85107415238&doi=10.1016%2fj.jeconom.2021.05.003&partnerID=40&md5=70ac5d23e604efd74bca505d7697cca3.

J. Paul Elhorst. Dynamic models in space and time. *Geographical Analysis*, 33(2):119 – 140, 2001. doi: 10.1111/j.1538-4632.2001.tb00440.x. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-0035000914&doi=10.1111%2fj.1538-4632.2001.tb00440.x&partnerID=40&md5=6f94b824952071aade1c226600e02f03.

J. Paul Elhorst. Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis*, 5(1):9 – 28, 2010. doi: 10.1080/17421770903541772. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-77951588750&doi=10.1080%2f17421770903541772&partnerID=40&md5=989ce1cd6549a430702052c8cc8b517b.

J. Paul Elhorst. Raising the bar in spatial economic analysis: two laws of spatial economic modelling. *Spatial Economic Analysis*, 19(2):115 – 132, 2024. doi: 10.1080/17421772.2024.2334845. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85190096198&doi=10.1080%2f17421772.2024.2334845&partnerID=40&md5=1c88b1bef5eea01f4c36ef08ad52200f.

J. Paul Elhorst, Marco Gross, and Eugen Tereanu. Cross-sectional dependence and spillovers in space and time: Where spatial econometrics and global var models meet. *Journal of Economic Surveys*, 35(1):192 – 226, 2021. doi: 10.1111/joes.12391. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090785936&doi=10.1111%2fjoes.12391&partnerID=40&md5=9ce2d6a3db1738931eeaa6eae6343a3d.

Michael F. Goodchild. The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers*, 94(2):300 – 303, 2004. doi: 10.1111/j.1467-8306.2004.09402008.x. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-28444486932&doi=10.1111%2fj.1467-8306.2004.09402008.x&partnerID=40&md5=d605533b3f672a5caa262fde7991da5f.

Siem Jan Koopman, Julia Schaumburg, and Quint Wiersma. Joint modelling and estimation of global and local cross-sectional dependence in large panels, 2021.

Junho Lee, Ronald E. Gangnon, and Jun Zhu. Cluster detection of spatial regression coefficients. *Statistics in Medicine*, 36(7):1118 – 1133, 2017. doi: 10.1002/sim.7172. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85005950024&doi=10.1002%2fsim.7172&partnerID=40&md5=01e6e82a2e07b86cfbd947110083d860.

Lung-Fei Lee and Jihai Yu. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154(2):165–185, 2010a. doi: doi.org/10.1016/j.jeconom.2009.08.001.

Lung-Fei Lee and Jihai Yu. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory*, 26(2):564 – 597, 2010b. doi: 10.1017/S0266466609100099. URL https://www.scopus.com/inward/

record.uri?eid=2-s2.0-77951095647&doi=10.1017%2fS0266466609100099&partnerID=40&md5=df080f13e8236cce04927f6d6c248776.

James LeSage and R. Kelley Pace. *Introduction to spatial econometrics*. 2009. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045659213&partnerID=40&md5=f36c5e559bdadaa988c0541a6335d735.

James P. LeSage. An introduction to spatial econometrics. *Revue d'Economie Industrielle*, 123(3):19 – 44, 2008. doi: 10.4000/rei.3887. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-84975123004&doi=10.4000%2frei.3887&partnerID=40&md5=99d9b43555794e6bf01be1a88a5dba7a.

Caterina Morelli, Paolo Maranzano, and Philipp Otto. Spatiotemporal clustering of GHGs emissions in Europe: exploring the role of spatial component, 2025. URL https://arxiv.org/abs/2503.11909. arXiv 2503.11909.

Cerqueti R., Maranzano P., and Mattera R. Spatially-clustered spatial autoregressive models with application to agricultural market concentration in Europe. *Journal of Agricultural, Biological, and Environmental Statistics*, 2025. doi: 10.1007/s13253-024-00672-4. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85214109557&doi=10.1007%2fs13253-024-00672-4&partnerID=40&md5=2ad5e0e9426f90a071aa5f42d5c29850.

Rozeta Simonovska. Sdpdmod: An r package for spatial dynamic panel data modeling. *Computational Economics*, 2025. doi: 10.1007/s10614-025-11056-2.

Liangjun Su, Wuyi Wang, and Xingbai Xu. Identifying latent group structures in spatial dynamic panels. *Journal of Econometrics*, 235(2):1955 – 1980, 2023. doi: 10.1016/j.jeconom.2023.02.007. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85150284978&doi=10.1016%2fj.jeconom.2023.02.007&partnerID=40&md5=9b64906a16238a2c87132760a208cb52.

Shonosuke Sugasawa and Daisuke Murakami. Spatially clustered regression. *Spatial Statistics*, 44, 2021. doi: 10.1016/j.spasta.2021.100525. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85108427980&doi=10.1016%2fj.spasta.2021.100525&partnerID=40&md5=820bf11087696d80f2c110544875cb9f.

W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240, 1970.

Jihai Yu, Robert de Jong, and Lung fei Lee. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. *Journal of Econometrics*, 146(1):118–134, 2008. doi: doi.org/10.1016/j.jeconom.2008.08.002.

A-Xing Zhu and Matthew Turner. How is the third law of geography different? *Annals of GIS*, 28(1):57 – 67, 2022. doi: 10.1080/19475683.2022.2026467. URL https://www.scopus.com/inward/record.uri?eid=2-s2.

0-85122804468&doi=10.1080%2f19475683.2022.2026467&partnerID=40&md5=f501b5d5a40bde7806736b8d21c4f98e.

A-Xing Zhu, Guonian Lu, Jing Liu, Cheng-Zhi Qin, and Chenghu Zhou. Spatial prediction based on third law of geography. *Annals of GIS*, 24(4):225 – 240, 2018. doi: 10.1080/19475683.2018.1534890. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85055540970&doi=10.1080%2f19475683.2018.1534890&partnerID=40&md5=baa1b241d70d7124d28337cf289e60ba.