The spatial evolution of economic activities: from theory to estimation

Davide Fiaschi [*]   Angela Parenti[†]   Cristiano Ricci[‡]

February 28, 2025

For a full version including Appenix see https://arxiv.org/pdf/2407.14267

Abstract

This paper studies the evolution of economic activities using a continuous time-space aggregation-diffusion model, which encompasses competing effects of agglomeration and congestion. To bring the model to the real data, a novel discretization technique over time and space is introduced. This technique effectively disentangles spatial effects into pure topography, agglomeration, repulsion, and diffusion forces, which is crucial for developing robust econometric methods in spatial economics. Our empirical analysis of personal income across Italian municipalities from 2008 to 2019 validates the model's primary predictions and demonstrates superior performance compared to the most common spatial econometric models in the literature.

1 Introduction

The investigation of the causes of the spatial distribution of economic activities is a very active field of analysis (Allen and Arkolakis, 2014; Redding and Rossi-Hansberg, 2017; Desmet et al., 2018). Several models have been introduced to

---

[*]University of Pisa, Via Ridolfi, 10, 56124 Pisa, Italy (e-mail: davide.fiaschi@unipi.it).
[†]University of Pisa, Via Ridolfi, 10, 56124 Pisa, Italy (e-mail: angela.parenti@unipi.it).
[‡]University of Pisa, Via Ridolfi, 10, 56124 Pisa, Italy (e-mail: cristiano.ricci@ec.unipi.it).

explain the emergence of spatial patterns characterized by geographical aggregations of economic activities, e.g. cities, and specific relationships among the locations of these aggregations, i.e. city size distribution, and their spatial organization (Krugman, 1991), as well as to quantify the effect of spatial agglomeration on aggregate income (Duranton and Puga, 2023). Despite this increasing literature, a definitive explanation of how aggregation economies emerge from the behaviour of individual agents is still needed (Rossi-Hansberg, 2019), and, such micro-foundations appear still more crucial in the light of the increasing availability of accurate data at fine geographical scale (Allen and Arkolakis, 2014; Ahlfeldt et al., 2015; Desmet et al., 2018). Furthermore, the complex nature of the spatial distribution of cross-sectional economic activities, together with its nonlinear evolution through time, needs specific quantitative techniques. Currently, neither spatial econometrics, which aims at capturing the spatial correlation in the *equilibrium outcome* from models of social/spatial interactions (e.g., Anselin, 2002; Brueckner, 2003; Ertur and Koch, 2007; Combes and Gobillon, 2015, and Xu and Lee, 2019), nor geo-statistical models, which focus on the specific spatial processes underlying the observed spatial distributions abstracting from the spatial interactions (e.g., Cressie and Wikle, 2015), are able to provide such techniques.

In this paper, we study the dynamics over time and space of economic activities by a micro-founded continuous time-space aggregation-diffusion model, where the collective macroscopic behaviour emerges from the dynamic of interacting and locally optimizing agents, in the mean-field limit as the number of agents becomes infinite. The macroeconomic dynamic is expressed by a Partial Differential Equation (PDE), which belongs to the class of *Aggregation-Diffusion Equations* (ADEs), which over the past 20 years have been employed in several biological applications and stimulated many mathematical works (see Carrillo et al., 2019a for a review). The competing effects of aggregation, repulsion and diffusion lead to several interesting properties, such as metastability, symmetrization, and non-uniqueness in the equilibrium solutions.[1] The estimation of ADEs is primarily addressed in two key works: Huang et al. (2019) tackle the challenge by observing individuals rather than aggregate variables, while Carrillo et al. (2024) focus specifically on estimating the interaction kernel, neglecting the broader parameter set. We propose a new discretization technique over time and space (Donaghy,

---

[1]A complete characterization of equilibria is not still available, and so far the only remarkable result is by Carrillo et al. (2019b), who proved that equilibria are radially symmetric in the absence of exogenous factors; for a specific class of interactions, the authors also proved the uniqueness of the equilibrium as well as the convergence to the unique equilibrium independent of the initial distribution.

2001; Aït-Sahalia, 2002; Piras et al., 2007; Oud et al., 2012; Wymer, 2012), which complements the literature on quantitative spatial economics (Redding and Rossi-Hansberg, 2017), whose outcome is a new class of spatial econometric models, denoted Spatial Aggregation Repulsion Diffusion (SARD) models. In particular, the use of *generalised finite difference methods* to transpose the system of PDEs defined in the continuum into a set of discrete locations (Jensen, 1972) allows for dealing with any non-uniform spatial organization of economic activities (and other socio-economic and cross-sectional variables), which are in general not arranged on a regular Cartesian grid but into irregular units, such as administrative regions.[2] Moreover, the formalism of PDEs in continuous space provides an innovative tool to introduce new elements in the theory of spatial econometrics. In particular, as described in more detail in Section 2, it allows to construct regressors having the remarkable feature of being cross-sectional zero-sum, which naturally leads to a disentangle between growth over time and spatial reallocation of economic activities. Here spatial reallocation is intended as a purely redistributive process by which if a location loses (gains) a certain amount, the same amount is distributed to (taken from) some other locations. Moreover, instead of simply measuring the effect of spatial correlation through local averages, by exploiting the differentials in the variables of interest across neighbouring locations, our approach is able to distinguish among the *aggregative* dynamics towards more appealing locations (a centripetal force, for example arising from the higher wages in specific labour markets), the *repulsive* dynamics (a centrifugal force, for example caused by the higher cost of living in a congested location), and the *diffusive* dynamics (an equalizing force, for example, caused by the random preferences of the underlying individuals).[3] Finally, the linear econometric model resulting from the discretization allows to directly study the process in the transient regime, i.e. not assuming to be in a neighbour of the equilibrium, therefore without relying on a (log-)linearization around the steady state (Ertur and Koch, 2007).

To summarise, relative to the existing literature on spatial econometrics and quantitative spatial economics, we propose a model which is able to disentangle growth over time from spatial reallocation and, for the latter, to separately identify topographical, aggregative, repulsive and diffusive forces. Numerical investigations show the capacity of our proposed procedure to correctly estimate the model's

---

[2]This constitutes one of the main differences with respect to spatial econometrics, where locations are assumed to be discrete by the beginning of analysis (Anselin, 2002).

[3]Here and in the rest of the paper the term *diffusion* is intended as in the literature of PDEs, that is a pure reallocation force which has the effect of driving the spatial process to a uniform distribution.

parameters when discretization over time and space is sufficiently fine. Maximum likelihood estimation appears to be the best method, although the simplest OLS produces in some circumstances acceptable outcomes. Moreover, our procedure is able to deal with the spatial correlation in the errors induced, among other factors, by the discretization.

An application to the spatio-temporal dynamics of Italian municipal income over the period 2008-2019 finds evidence of the presence of agglomerative, repulsive and diffusive forces, jointly with the key role of geographical characteristics. SARD model outperforms the most common spatial econometric models used in the literature, as spatial LAG and DURBIN model (LeSage and Pace, 2009). We also show the capacity of the SARD model to quantify the contribution of each reallocation force to the average growth rate of each municipality, as well as to the overall (di)convergence in local incomes.

There are two important limitations in our analysis. First, the SARD model can be thought as a reduced form of a structural spatial model of income (see Fiaschi and Ricci, 2023), which by its nature limits the possibility to be used to explore the dynamics of other key economic variables such as local wages, prices, rents, etc. It remains that it can be used both for forecast as well as policy purposes in the same vein as the VAR approach. Second, our methodology requires a sufficiently high level of geographical resolution in order to prevent significant discretization bias in the estimation.

The paper is organized as follows: Section 2 presents a prototype model of spatial growth; Section 3 discusses the methodology to estimate such a model and explore its properties by numerical investigations; Section 4 discusses an empirical application of the proposed methodology to Italian municipalities; and, finally, Section 5 concludes. Technical details are gathered in the Appendix.

2   A prototype model of spatial growth

This section introduces a class of spatial growth models encompassing the main features present in the literature, i.e. i) the spatial non-uniformity caused by geographical and socio-economic factors; ii) the spatial aggregation of economic activities driven by positive spatial spillovers; iii) the centrifugal dynamics driven by congestion and excessive crowding; and, finally, iv) the existence of randomness in the individuals' preferences (Fujita and Thisse, 2002).

Let $y(t, z)$ be the variable of interest of our model in location z at time t, e.g. municipal total income per square kilometre, where $\Omega \subseteq \mathbb{R}^2$. Each point $z \in \Omega$

is identified by two components $z = (z_1, z_2)$, i.e. latitude and longitude. The dynamics of $y(t, z)$ is assumed to obey the following partial differential equation (PDE):

$$
\begin{aligned}
\partial_t y(t, z) &= a(t, z) + \varphi y(t, z) + \\
&+ \gamma_S \text{div}_z \left(y(t, z) \nabla_z S(z) \right) + \\
&+ \gamma_A \text{div}_z \left(y(t, z) \nabla_z \left(K_{h_A} * y \right)(t, z) \right) + \\
&+ \gamma_R \text{div}_z \left(y(t, z) \nabla_z \left(K_{h_R} * y \right)(t, z) \right) + \\
&+ \gamma_D \Delta_z y(t, z),
\end{aligned}
\tag{1}
$$

with $\gamma_D \geq 0$.[4] Eq. (1) makes wide use of concepts used in PDE literature, which will be discussed in detail below. In particular, the variable on the left-hand side $\partial_t y(t, z)$ represents the total variation of the quantity $y(t, z)$ at time t in location z, expressed through the use of the partial derivative ∂_t with respect to time. In particular, div_z is the *divergence operator*, i.e. $\text{div}_z F \equiv \partial_{z_1} F_{z_1} + \partial_{z_2} F_{z_2}$, where $F = (F_{z_1}, F_{z_2})$ is a vector field $F : \Omega \to \mathbb{R}^2$ and ∂_{z_i} are the partial derivatives with respect to each of the components of $z = (z_1, z_2)$. All the partial derivatives expressed above are to be intended in the spatial sense, i.e. they measure how the spatial profile varies along one of the two components, either latitude or longitude. When $f : \Omega \to \mathbb{R}$ is a scalar function the term $\nabla_z f$ stands for the *gradient operator*, i.e. $\nabla_z f \equiv (\partial_{z_1} f, \partial_{z_2} f)$ while finally the term $\Delta_z f \equiv \partial^2_{z_1 z_1} f + \partial^2_{z_2 z_2} f$ is called the *Laplace operator* and involves the second order pure partial derivatives $\partial^2_{z_1 z_1}$ and $\partial^2_{z_2 z_2}$.

Neglecting boundary effects one has

$$
\int_\Omega \text{div}_z \left(y(t, z) \nabla_z S(z) \right) dz = 0;
$$

$$
\int_\Omega \text{div}_z \left(y(t, z) \nabla_z \left(K_{h_A} * y \right)(t, z) \right) dz = 0;
$$

$$
\int_\Omega \text{div}_z \left(y(t, z) \nabla_z \left(K_{h_R} * y \right)(t, z) \right) dz = 0; \text{ and}
$$

$$
\int_\Omega \Delta_z y(t, z) dz = 0,
$$

which expresses that the last four terms of Eq. (1) are purely reallocative.

Fiaschi and Ricci (2023) shows that Eq. (1) can be derived in the limit of an infinity-agent economy, where agents can be understood as units of production/-

---

[4]Eq. (1) is general not well-posed, i.e solutions often grow unbounded in finite time or even fail to exist, in the case $\gamma_D < 0$, see Taylor (1975).

consumers, as the outcome of the agents' mobility driven by locally maximizing behaviour (Appendix C summarise the key steps in the derivation).

The first term on the right-hand side of Eq. (1) $a(t,z)$ accounts for the *exogenous local component* of the dynamics, e.g. migration shocks, as well as *local effects* (amenities) and *time effects* (e.g., exogenous technological progress, see Ertur and Koch, 2007).

Parameter φ instead represents the change in $y(t,z)$ due to the *local specific process of accumulation*, i.e. the impact of variable $y(t,z)$ on the time change of the variable itself, independent of any spatial interaction. Taking as reference the Solovian framework, φ should reflect the *shape of production function, saving behaviour, depreciation rates of factors* and *employment growth*. The dynamic at the aggregate level is given by:

$$\int_\Omega \partial_t y(t,z)dz = \dot{Y}(t) = \int_\Omega a(t,z)dz + \varphi Y(t), \tag{2}$$

which points out as φ represents the part of the *aggregate long-run growth rate of the economy* independent of the spatial distribution of activities, while the first term in Eq. (2) represents the aggregate effect of exogenous growth components.

The main reference behind the terms from the second to the fifth line of the right-hand side of Eq. (1) is the so-called *Fokker-Planck Equation*. In particular, they contain *differential operators*, which belong to the language of PDEs, and are particularly effective for describing the various sources of change in the spatial distribution of $y(t,z)$. Since partial derivatives express a measure of differentials across different locations, the sign of their coefficients reflects the direction driving the reallocation. For example, as the second line, if the coefficient γ_S is positive, then the term $\gamma_S \mathrm{div}_z(y(t,z)\nabla_z S(z))$ expresses the tendency of $y(t,z)$ to increase in those locations where $S(z)$ is lower, and to decrease of the same amount where $S(z)$ is higher.[5] The definition of higher and lower is to be intended as a relative comparison between locations and is provided by the use of $\nabla_z S(z)$ which measures the *steepness of the transition* moving from one location to the other. This increase-decrease effect is made in such a way that the total amount of mass in the distribution of y which is gained where it increases, is lost in the other locations.

---

[5] A simple example in one dimension showing the interpretation of $\gamma_S > 0$ is the following. Assume that the distribution $y(z)$ is uniform in the space, i.e $y(z) = \overline{y} \; \forall z$, and the function $S(z)$ is a sigmoid of the form $S(z) = \frac{1}{1+e^{-z}}$, i.e. increasing in z and with $S''(z) < 0$ for $z > 0$. In this case, the second term of Eq. (1) becomes $\gamma_S \overline{y} S''(z)$ and, therefore: i) for $z > 0$, $\gamma_S \overline{y} S''(z)$ is negative, implying that the values of $y(z)$ are decreasing; while, ii) for $z < 0$, $\gamma_S \overline{y} S''(z)$ is positive, implying that the values of $y(z)$ are increasing.

Hence, the total variation accounted for by the second term, when one takes into account all the locations, is zero. For this reason, we refer to all the terms in Eq. (1) except for the first two, as purely *reallocation* terms, since they don't affect the total amount of y.

The second line of Eq. (1) $\gamma_S \text{div}_z (y(t,z)\nabla_z S(z))$ is introduced to take into account the *topography* of Ω. In particular, $\nabla_z S(z)$ indicates the *pure geographical and exogenous* (dis)advantages to move from location z to neighbouring locations. In this regard, the coefficient γ_S is expected to be positive reflecting a reallocation towards places where the exogenous variable $S(z)$ reaches its lowest levels (e.g. from the mountains towards the plains).

The third line on the right-hand side of Eq. (1) $\gamma_A \text{div}_z (y(t,z)\nabla_z (K_{h_A} * y)(t,z))$ represents the effect of *aggregation* and of cultural features such as roads, land boundaries, and buildings, i.e. the tendency of y to concentrate in specific locations. Fiaschi and Ricci (2023) shows that this term, and the next two, of Eq. (1) are the outcome of the agents' mobility driven by locally maximizing behaviour. The intensity of this process is measured by $\gamma_A < 0$. K_{h_A} is a kernel function with bandwidth $h_A \geq 0$, and "$*$" is the *convolution operator*, i.e. $(K_{h_A} * y)(t,z) \equiv \int_\Omega K_{h_A}(k-z) y(t,k)dk$. In other words, $(K_{h_A} * y)(t,z)$ is the weighted sum of all y around location z at period t, where the weights are defined by kernel K_{h_A}; in particular, the shape of K_{h_A} and the value of h_A decide how these weights change with the distance from location z. Since $\nabla_z (K_{h_A} * y)(t,z)$ points to the directions where the level of $y(t,z)$ is higher when averaged among neighbours, it plays a similar role of $\nabla_z S(z)$ as the term on the second line of Eq. (1). An important difference is that, instead of being purely exogenous as for $S(z)$, the direction which drives the reallocation is now endogenous since it depends on the spatial distribution of y. In particular, we assume $\gamma_A < 0$, so that the reallocation is driven towards those areas where the local average of y, $(K_{h_A} * y)(t,z)$ is higher, providing the intuition on why this term will tend to concentrate y over space. A standard explanation in economics is the observed process of aggregation of workers, i.e. the emergence of cities based on the positive externalities generated by working in places where other activities and/or skilled workers are already present (Fujita and Thisse, 1996; Krugman, 1998; Moretti, 2004; Ahlfeldt et al., 2015).

The fourth line on the right-hand side of Eq. (1) $\gamma_R \text{div}_z (y(t,z)\nabla_z (K_{h_R} * y)(t,z))$ represents the effect of *repulsion* of y across different locations, that is the tendency of y to flow away from locations with higher levels. The intensity of this process is measured by $\gamma_R > 0$. This term is exactly analogous to the one related to aggregation, except that the Kernel function K_{h_R} can be different from K_{h_A} (for

example, in the speed of decay as a function of the distance), and that the sign γ_R is assumed to be positive, so the effect instead of being centripetal is centrifugal. The population and income density, increasing the price of lands and living, operates as a centrifugal force (Duranton and Puga, 2020). For Kernels K_{h_A} and K_{h_R} we adopt distance-decay functions as:

$$K_{h_A}(z) \equiv \begin{cases} \left(\frac{c_K}{h_A}\right)^2 \frac{1}{\left(\left\|\frac{z}{h_A}\right\|+1\right)^2} & \text{if } \|z\| \leq h_A; \\ 0 & \text{otherwise} \end{cases},$$

$$K_{h_R}(z) \equiv \begin{cases} \left(\frac{c_K}{h_R}\right)^2 \frac{1}{\left(\left\|\frac{z}{h_R}\right\|+1\right)^2} & \text{if } \|z\| \leq h_R; \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

where $c_K = 1/\sqrt{2\pi (\log 2 - 1/2)}$, h_A and h_R are the distance within which there exists a non-null (aggregative and repulsive) spatial interaction.

Finally, the last line on the right-hand side of Eq. (1) $\gamma_D \Delta_z y(t,z)$, represents the effect of *diffusion* across different locations of y, which tends to uniformly spread y over space. The intensity of this diffusion process in location z at time t depends on the parameter $\gamma_D > 0$ and on the sign and magnitude of second derivatives of $y(t,z)$.[6] In economics, the diffusion process can be justified by the *idiosyncratic random component* in the agents' choice, determined by their hidden characteristics/preferences (Wozniak, 2010). At the aggregate level, this idiosyncratic component drives the distribution of the variable of interest in space towards a uniform distribution.

While is it true that both *repulsion* and *diffusion* represent centrifugal forces, they are very different in nature. The repulsion effect only takes place in the presence of overcrowding; the diffusion, instead, is always affecting the dynamics. In particular, diffusion always tends to equalize all the levels of $y(t,z)$ across locations, while repulsion only lowers the level of y with high density. In other words, in the same way that aggregation expresses the tendency of individuals to relocate to be closer to each other, repulsion can be seen as an effect due to *congestion* and *crowding*. A possible source of the observed outflows of individuals

---

[6]The use of the second derivative can be understood intuitively if one thinks of the one-dimensional example of a bell-shaped distribution of $y(t,z)$ (e.g. a Gaussian distribution). In this case, the spatial second derivative measures the convexity/concavity of the distribution, being negative in the centre of the peak (concave) and positive in the area outside the peak (convex). Therefore, when $\gamma_D > 0$ the effect is to decrease where the distribution is convex (i.e. on the local peaks) and to increase in all other regions. Farlow (1993, p. 12) provides an intuition of why the second derivative is crucial for describing a diffusion process, which tends to uniformly spread the variable of interest over space.

from very crowded locations can arise from the higher housing prices and moving costs and, in general, from the higher cost of living in locations with high population density (Krugman, 1998). Moreover, this tendency is generally justified by the factor-return equalization across different locations in the presence of decreasing marginal returns to factors.[7] Finally, theory suggests that some signs restrictions have to be tested in the estimation; in particular, in Eq. (1) $\gamma_S \leq 0, \gamma_A \leq 0$, to reflect the reallocation due to topography and the aggregation effect respectively, while $\gamma_R \geq 0, \gamma_D \geq 0$ to reflect the repulsive effect and the presence of diffusion.

3   From theory to empirics

This section discusses a methodology to estimate the parameters of Eq. (1) when only a sample of N spatial units observed at times 0 and τ is available. In particular, Section 3.1 reports the main steps of the estimate; Section 3.2 describes in detail how the econometric model used in the estimate is derived by an appropriate approximation over time and space of Eq. (1); Section 3.3 contains some remarks on the derivation of the econometric model and its possible extensions; Section 3.4 discusses some possible issues in the estimate; finally, Section 3.5 investigates the properties of the proposed methodology by numerical simulations.

3.1   The econometric model

The econometric model associated to Eq. (1), denoted *Spatial Aggregation Repulsion Diffusion* (SARD), is given by:

$$
\begin{aligned}
\Delta_\tau \mathbf{y} &= \tilde{\alpha}\mathbf{1} + \tilde{\varphi}\mathbf{y} + \tilde{\gamma}_S \mathbf{x}_S + \tilde{\gamma}_A \mathbf{x}_A + \tilde{\gamma}_R \mathbf{x}_R + \tilde{\gamma}_D \mathbf{x}_D + \\
&+ \tilde{\rho}_S M_S \Delta_\tau \mathbf{y} + \tilde{\rho}_A M_A \Delta_\tau \mathbf{y} + \tilde{\rho}_R M_R \Delta_\tau \mathbf{y} + \tilde{\rho}_D M_D \Delta_\tau \mathbf{y} + \\
&+ \boldsymbol{\epsilon},
\end{aligned}
\tag{4}
$$

where $\Delta_\tau \mathbf{y}$ is the vector of the time changes of $\mathbf{y}$ of length N (the number of the units of observations), $\mathbf{1}, \mathbf{y}, \mathbf{x}_S, \mathbf{x}_A, \mathbf{x}_R, \mathbf{x}_D$ are the vectors of regressors of length N, M_S, M_A, M_R, M_D, are $N \times N$ matrices whose calculation is described

---

[7]Suppose to consider two locations, 1 and 2, with a different endowment of capital $k_1 > k_2$ but with the same production function; then $f'(k_1) < f'(k_2)$ under the hypothesis of $f''(\cdot) < 0$; with free movement of capital we should observe a flow of capital from location 1 to location 2. The second derivative with respect to the distribution over space of capital is a proxy for the difference in the level of k_1 and k_2, which, in turn, is reflected in the difference in factor returns and, hence, in the intensity of reallocation.

in Section 3.2, and $\boldsymbol{\epsilon}$ is an error component whose characteristics are discussed below.

Section 3.2 shows that Model (4) represents a (corrected) first order Taylor approximation over time and a second order approximation over space of Eq. (1). However, the parameters $\tilde{\alpha}$, $\tilde{\varphi}$, $\tilde{\gamma}_S$, $\tilde{\gamma}_A$, $\tilde{\gamma}_R$, $\tilde{\gamma}_D$ correspond to those of the theoretical model up to a scale factor, i.e:

$$\tilde{\alpha} = \frac{\alpha}{1 - \tau \rho_\varphi/2}, \quad \tilde{\varphi} = \frac{\varphi}{1 - \tau \rho_\varphi/2}, \quad \tilde{\gamma}_j = \frac{\gamma_j}{1 - \tau \rho_\varphi/2}, \quad \tilde{\rho}_j = \frac{\tau \rho_j/2}{1 - \tau \rho_\varphi/2}, \quad (5)$$

with $j \in \{S, A, R, D\}$. To identify parameters of Eq. (1) we use the dynamics of aggregate variable of interest Y from Eq. (2), i.e.:

$$Y(\tau) = \left(Y(0) + \frac{\alpha}{\varphi}\right) \exp(\varphi \tau) - \frac{\alpha}{\varphi}, \quad (6)$$

to get an estimate of ρ_φ, i.e.[8]

$$\rho_\varphi = \left(\frac{2}{\tau}\right) \left[1 - \left(\frac{1}{\tilde{\varphi}\tau}\right) \log\left(\frac{Y(\tau) + \tilde{\alpha}/\tilde{\varphi}}{Y(0) + \tilde{\alpha}/\tilde{\varphi}}\right)\right]. \quad (7)$$

Model (4) admits a panel structure with individual and time fixed effects as shown in Section 3.2. In particular, assuming that $a(t, z) = \alpha(z) + d(t)$, then:

$$\tilde{\boldsymbol{\alpha}} \equiv \frac{\boldsymbol{\alpha}}{1 - \tau \rho_\varphi/2} \quad \text{and} \quad \tilde{\mathbf{d}} \equiv \frac{\mathbf{d} + \tau \Delta_\tau \mathbf{d}/2}{1 - \tau \rho_\varphi/2}, \quad (8)$$

where $\tilde{\boldsymbol{\alpha}}$ is the vector of location fixed effects of length N and $\tilde{\mathbf{d}}$ is the vector of time fixed effect of length equal to the number of time periods. The identification of $\tilde{\boldsymbol{\alpha}}$ and $\tilde{\mathbf{d}}$ however requires a sample of more than two periods and to specify a functional form for $d(t)$ (e.g. a linear trend) in order to identify the parameters by exploiting the aggregate dynamics of Y given by:

$$Y(t) = \left(Y(t_0) + \frac{\bar{\alpha}}{\varphi}\right) \exp(\varphi(t - t_0)) - \frac{\bar{\alpha}}{\varphi} + \exp(\varphi(t - t_0)) \int_{t_0}^t \exp(-\varphi(s - t_0)) d(s) ds, \quad (9)$$

for $t_0 \in \{0, \tau, 2\tau, \cdots\}$ and $t \in \{\tau, 2\tau, \cdots\}$, where $\bar{\alpha} \equiv \int_\Omega \alpha(z) \, dz$.

The error term $\boldsymbol{\epsilon}$ in the Model (4) includes, in addition to the classical i.i.d stochastic component, the randomness deriving from the finite number of individuals and firms and the potential presence of measurement errors. In particular, assuming

---

[8]Substituting the expression for α and φ from Eq. (5) into Eq. (6) we get an equation for ρ_φ.

$y_{it}^{OBS} = y_{it} + \xi_{it}$ for $t \in \{0, \tau\}$, where y_{it}^{OBS} is the *observed* value of the variable of interest and ξ_{it} is a classical error (i.e. a random variable i.i.d. over time and space), $\partial_t y_{it}^{OBS} = \partial_t y_{it} + d\xi_{it}$, i.e. the dynamics of the observed variable is the sum of two components: the first one systematic, which induces a structure of spatial correlation (see Section 3.2.3), while the second one purely random. Inspired by the spatial econometric literature (LeSage and Pace, 2009, p. 27), we consider:

$$
\begin{aligned}
\boldsymbol{\epsilon} &= \lambda W_\epsilon \boldsymbol{\epsilon} + \boldsymbol{\eta}, \text{ and} \\
\boldsymbol{\eta} &\sim \mathcal{N}(0, \sigma^2 I),
\end{aligned}
\tag{10}
$$

where λ is a parameter measuring the intensity of spatial spillovers in the error component, W_ϵ is a $N \times N$ spatial weights matrix, and σ^2 is the variance of the i.i.d. random component. The choice of W_ϵ is discussed in Section 3.2 and Appendix B. It is worth noticing that the specification of the error in Eq. (10) as a spatial error only affects the efficiency of estimates but not their consistency (LeSage and Pace, 2009, p. 28).

From the theoretical model, we expected the following signs for the estimated parameters: i) $\varphi(\tilde{\varphi}) > 0$, which should reflect the positive impact of initial conditions on the variation of the variable of interest over time; ii) $\gamma_S(\tilde{\gamma}_S) < 0$ representing the spatial effect of reallocating towards more spatially appealing locations; iii) $\gamma_A(\tilde{\gamma}_A) < 0$ indicating evidence of aggregation effect over space; iv) $\gamma_R(\tilde{\gamma}_R) > 0$ showing evidence of repulsion effect; finally v) $\gamma_D(\tilde{\gamma}_D) > 0$, highlighting a diffusion effect over space.

3.2 The derivation of SARD model

In this section, we describe in detail the procedure to derive Model (4) in the presence of N units of observation z^i for $i = 1, \ldots N$ on $T+1$ periods $t = 0, \ldots, T$. Let

$$
\int_{t-1}^{t} y(s, z^i) \, ds
$$

be the total cumulative value of the variable of interest in *period* t in unit z^i. Since the variable of interest is observed only on periods $t = 0, \ldots, T$ we assume that the flow $y(s, z^i)$ is homogeneous for $s \in [t-1, t]$, and therefore we consider

$$
y_{ti} \equiv y(t, z^i),
\tag{11}
$$

the discrete measure of our variable of interest at period t in spatial unit z^i.[9] With $y_i \equiv y(0, z^i)$ we denote its measure at period 0.

3.2.1 Discretization over space

As the discretization over space, starting from the second line of Eq. (1), $\gamma_S \text{div}_z \left(y(t,z) \nabla_z S(z) \right)$, making partial derivatives with respect to (z_1, z_2) explicit yields:

$$\gamma_S \text{div}_z \left(y(t,z) \nabla_z S(z) \right) = \gamma_S \left[\partial_{z_1} \left(y(t,z) \partial_{z_1} S(z) \right) + \partial_{z_2} \left(y(t,z) \partial_{z_2} S(z) \right) \right]. \quad (12)$$

Using the notation introduced in Appendix E, Matrices M_{z_1} and M_{z_2} respectively denote the (approximate) derivative operators of the first order partial derivatives ∂_{z_1} and ∂_{z_2}, i.e., for unit z^i:

$$\partial_{z_1} y(t,z) \big|_{z=z^i} \approx (M_{z_1} \mathbf{y}_t)_i \text{ and } \partial_{z_2} y(t,z) \big|_{z=z^i} \approx (M_{z_2} \mathbf{y}_t)_i, \quad (13)$$

where $\mathbf{y}_t = (y_{t1}, ..., y_{tN})$.

The smaller the distance among contiguous spatial units the more accurate the approximation in Eq. (13) (see Thomas, 2013 for a reference in the case of a uniform grid). From Eqq. (12) and (13) the discrete counterpart of the term in the second line in Eq. (1) is given by:

$$\gamma_S \text{div}_z \left(y(t,z) \nabla_z S(z) \right) \approx \gamma_S \left[M_{z_1} \left(\mathbf{y}_t \odot M_{z_1} \mathbf{s} \right) + M_{z_2} \left(\mathbf{y}_t \odot M_{z_2} \mathbf{s} \right) \right],$$

with $s_i \equiv S(z^i)$, $\mathbf{s} = (s_1, ..., s_N)$ and where $\odot$ denotes the element-wise product between vectors. The approximation of the third and fourth line of Eq. (1) follows in the same manner, but the discretization of the convolution terms $(K_{h_A} * y)(t, z)$ and $(K_{h_R} * y)(t, z)$. In particular, taking

$$(K_{h_A} * y)(t, z) \equiv \int_\Omega K_{h_A} (k - z) \, y(t, k) dk \approx W_{h_A} \mathbf{y}_t, \quad (14)$$

with W_{h_A} a $(N \times N)$ matrix representing the kernel K_{h_A} defined as:

$$(W_{h_A})_{i,j} = K_{h_A} \left(z^i - z^j \right) A_j, \quad (15)$$

and A_j the area of the j-th spatial unit, then the fourth term in Eq. (1) can be

---

[9]It is also possible to consider the case where the process is observed at dates $0, \Delta, 2\Delta, \ldots, T\Delta$ where $\Delta > 0$ is fixed. This amounts to consider the cumulative variation as $\int_{(t-1)\Delta}^{t\Delta} y(s, z^i) \, ds$, reducing Eq. (11) to $y_{ti} \equiv y(t\Delta, z^i) \cdot \Delta$. In the text, we have assumed $\Delta \equiv 1$ (i.e. 1 year in Section 4).

approximated as:

$$\gamma_A \text{div}_z \left(y(t,z) \nabla_z \left(K_{h_A} * y \right)(t,z) \right) \approx \gamma_A \left[M_{z_1} \left(\mathbf{y}_t \odot M_{z_1} W_{h_A} \mathbf{y}_t \right) + M_{z_2} \left(\mathbf{y}_t \odot M_{z_2} W_{h_A} \mathbf{y}_t \right) \right].$$
(16)

The same applies to the fifth term in Eq. (1), with W_{h_R} a $(N \times N)$ matrix representing the kernel K_{h_R}. W_{h_A} and W_{h_R} have a clear similarity with the spatial weight matrices used in the spatial econometric literature (LeSage and Pace, 2009); however, since the matrices W_{h_A} and W_{h_R} approximate in discrete space integrals in continuous space, the value of the function in the centre of the neighbour where integration takes place, i.e. their diagonal elements, are non-zero.

The (approximate) derivative operators of the second order partial derivatives $\partial^2_{z_1 z_1}$ and $\partial^2_{z_2 z_2}$ for unit z^i are

$$\partial^2_{z_1 z_1} y(t,z) \big|_{z=z^i} \approx (M_{z_1 z_1} \mathbf{y}_t)_i \text{ and } \partial^2_{z_2 z_2} y(t,z) \big|_{z=z^i} \approx (M_{z_2 z_2} \mathbf{y}_t)_i,$$
(17)

which can be used to approximate the sixth term of Eq. (1). The overall outcome of the discretization over space is therefore given by:

$$
\begin{aligned}
\partial_t y\left(t, z^i\right) & = a\left(t, z^i\right) + \varphi y_{ti} + \\
& + \gamma_S \left[M_{z_1} \left(\mathbf{y}_t \odot M_{z_1} \mathbf{s} \right) + M_{z_2} \left(\mathbf{y}_t \odot M_{z_2} \mathbf{s} \right) \right]_i + \\
& + \gamma_A \left[M_{z_1} \left(\mathbf{y}_t \odot M_{z_1} W_{h_A} \mathbf{y}_t \right) + M_{z_2} \left(\mathbf{y}_t \odot M_{z_2} W_{h_A} \mathbf{y}_t \right) \right]_i + \\
& + \gamma_R \left[M_{z_1} \left(\mathbf{y}_t \odot M_{z_1} W_{h_R} \mathbf{y}_t \right) + M_{z_2} \left(\mathbf{y}_t \odot M_{z_2} W_{h_R} \mathbf{y}_t \right) \right]_i + \\
& + \gamma_D \left[(M_{z_1 z_1} + M_{z_2 z_2}) \mathbf{y}_t \right]_i + \\
& + r_s^{ti},
\end{aligned}
$$
(18)

where r_s^{ti} is a reminder of the discretization over space such that $||\mathbf{r}_s^t||$ goes to zero as the space discretization gets finer. The statistical properties of this reminder will be explored in Section 3.5.

3.2.2 Discretization over time

Assume $\tau \in \mathbb{N}$ and $\tau > 1$. As the discretization in time, Taylor's theorem in Lagrange form states that there exists a $t' \in (0, \tau)$ such that:[10]

$$\partial_t y(0, z^i) = \Delta_\tau y_i - \frac{1}{2} \tau \partial^2_{tt} y(t', z^i),$$
(19)

---

[10]See Theorem 5.15 in Rudin et al. (1964).

where

$$\Delta_\tau y_i \equiv \frac{y(\tau, z^i) - y(0, z^i)}{\tau} \tag{20}$$

is the time variation of y in the period $[0, \tau]$, from which:

$$\Delta_\tau y_i = \partial_t y(0, z^i) + \frac{1}{2}\tau \partial_{tt}^2 y(t', z^i), \tag{21}$$

where from the previous, under the assumption that $\partial_{tt}^2 y$ is bounded, one can see that the discrete derivative is a good approximation of the "true" derivative as τ approaches zero. A "naive" discretization over time would consider only the first term on the right-hand side in Eq. (21), which can be directly taken from Eq. (1) evaluated in $t = 0$ using Eq. (18). The second term on the right-hand side of Eq. (21) measures the error in the discretization of considering the first term only. In our case, this error has a non negligible magnitude due to the time length of discretization, the exponential shape in the dynamics, and the presence of spatial spillovers. To illustrate the point, assume that $\gamma_S, \gamma_A, \gamma_R = 0$; from Eq. (1) for $t = t'$, it follows:

$$\begin{aligned}
\partial_{tt}^2 y(t', z^i) &= \partial_t \left[\partial_t y(t, z^i)\right]\big|_{t=t'} = \\
&= \partial_t a\left(t, z^i\right)\big|_{t=t'} + \varphi \partial_t y(t, z^i)\big|_{t=t'} + \gamma_D \left[\partial_{z_1 z_1} \partial_t y(t', z)\big|_{z=z^i} + \partial_{z_2 z_2} \partial_t y(t', z)\big|_{z=z^i}\right].
\end{aligned} \tag{22}$$

Since t' is not determined, a natural choice is to approximate the first order time derivative in unit z^i at period t' by the time variation in the whole period, i.e. $\partial_t y(t', z^i) \approx \Delta_\tau y_i$ and, at the same time, to run the space discretization as in Eq. (18). This approximation is more accurate as t' goes to t, i.e. τ goes to zero. Consequently, the coefficients φ and γ_D appearing in Eq. (22) will be affected by this approximation. Therefore:

$$\partial_{tt}^2 y(t', z^i) \approx \partial_t a\left(t', z^i\right) + \rho_\varphi \Delta_\tau y_i + \rho_D \left[(M_{z_1 z_1} + M_{z_2 z_2})\Delta_\tau \mathbf{y}\right]_i, \tag{23}$$

for some $\rho_\varphi \approx \varphi$ and $\rho_D \approx \gamma_D$. Hence, from Eq. (21):

$$\begin{aligned}
\Delta_\tau y_i &\approx a\left(0, z^i\right) + \varphi y_i + \gamma_D \left[(M_{z_1 z_1} + M_{z_2 z_2})\mathbf{y}\right]_i + \\
&+ \frac{\tau}{2}\left\{\partial_t a\left(t, z^i\right)\big|_{t=0} + \rho_\varphi \Delta_\tau y_i + \rho_D \left[(M_{z_1 z_1} + M_{z_2 z_2})\Delta_\tau \mathbf{y}\right]_i\right\},
\end{aligned}$$

or, taking $a_i \equiv a\left(0, z^i\right)$ and $\Delta_\tau a_i \approx \partial_t a\left(t, z^i\right)\big|_{t=0}$, in matrix form:

$$\begin{aligned}
\Delta_\tau \mathbf{y} &\approx \mathbf{a} + \varphi \mathbf{y} + \gamma_D (M_{z_1 z_1} + M_{z_2 z_2}) \mathbf{y} + \\
&\quad + \frac{\tau}{2} \left[\Delta_\tau \mathbf{a} + \rho_\varphi \Delta_\tau \mathbf{y} + \rho_D (M_{z_1 z_1} + M_{z_2 z_2}) \Delta_\tau \mathbf{y} \right],
\end{aligned}$$

from which:

$$\begin{aligned}
\Delta_\tau \mathbf{y} \approx \left[\left(1 - \frac{\tau \rho_\varphi}{2}\right) \mathbf{I} - \frac{\tau \rho_D}{2} (M_{z_1 z_1} + M_{z_2 z_2}) \right]^{-1} \times \\
\times \left[\mathbf{a} + \frac{\tau}{2} \Delta_\tau \mathbf{a} + \varphi \mathbf{y} + \gamma_D (M_{z_1 z_1} + M_{z_2 z_2}) \mathbf{y} \right]. \quad (24)
\end{aligned}$$

Eq. (24) makes clear the importance of considering the second term in Eq. (21), which is represented by the expression within the square bracket to be inverted. The magnitude of this correction is directly related to τ, i.e. the length of time discretization, and to the extent of spatial spillovers measured by matrices $M_{z_1 z_1}$ and $M_{z_2 z_2}$. Moreover, Eq. (24) shows that all the parameters are identifiable up to the scale factor $(1 - \tau \rho_\varphi / 2)$. Appendix A deals with the general case, and it shows that:

$$\begin{aligned}
\Delta_\tau \mathbf{y} &= \tilde{\mathbf{a}} + \tilde{\varphi} \mathbf{y} + \\
&\quad + \tilde{\gamma}_S \left[M_{z_1} (\mathbf{y} \odot M_{z_1} \mathbf{s}) + M_{z_2} (\mathbf{y} \odot M_{z_2} \mathbf{s}) \right] + \tilde{\rho}_S M_S \Delta_\tau \mathbf{y} \\
&\quad + \tilde{\gamma}_A \left[M_{z_1} (\mathbf{y} \odot M_{z_1} W_{h_A} \mathbf{y}) + M_{z_2} (\mathbf{y} \odot M_{z_2} W_{h_A} \mathbf{y}) \right] + \tilde{\rho}_A M_A \Delta_\tau \mathbf{y} \\
&\quad + \tilde{\gamma}_R \left[M_{z_1} (\mathbf{y} \odot M_{z_1} W_{h_R} \mathbf{y}) + M_{z_2} (\mathbf{y} \odot M_{z_2} W_{h_R} \mathbf{y}) \right] + \tilde{\rho}_R M_R \Delta_\tau \mathbf{y} \\
&\quad + \tilde{\gamma}_D (M_{z_1 z_1} + M_{z_2 z_2}) \mathbf{y} + \tilde{\rho}_D M_D \Delta_\tau \mathbf{y} \\
&\quad + \mathbf{r}_{s,t}, \quad (25)
\end{aligned}$$

where M_S, M_A, M_R and M_D are defined in Appendix A and $\mathbf{r}_{s,t}$ is the reminder of the process of discretization over space and time and it is such that $||\mathbf{r}_{s,t}||$ goes to zero as the space and time discretization gets finer. The statistical properties of this reminder will be explored in Section 3.5.

3.2.3 The reminder of the discretization as an error term

Some of the properties of the reminder $\mathbf{r}_{s,t}$ of Model (25) are still unknown. This suggests to treat it as an error term in the estimate. Our guess is that such a term can show some spatial dependence induced by the spatial discretization. Moreover, Eq. (1) is derived by considering an infinite number of agents in every location $z \in \Omega$ in the spirit of the laws of large numbers (see Appendix C). However, in the

15

real world the units of observation can include many agents but not infinite and, therefore, some fluctuations from the mean behaviour should be present at the unit level. The characteristics of these fluctuations are still under investigation on the theoretical ground (Graham et al., 1996). Heuristically, the presence of spatial interactions between agents suggests that fluctuations should contain some degree of spatial correlation already in the continuous time-space framework. This spatial correlation should be further magnified in the discretization over time and space by the presence of spatial spillovers in the dynamics (represented by matrices M_S, M_A, M_R and M_D, and W_{h_R} and W_{h_A}).

To account for the reminders of the approximation over time and space and of the possible finite-world fluctuations, Model (25) is therefore amended as it follows:

$$
\begin{aligned}
\Delta_\tau \mathbf{y} \ = \ & \tilde{\mathbf{a}} + \tilde{\varphi}\mathbf{y} + \\
& + \ \tilde{\gamma}_S \underbrace{\left[M_{z_1}\left(\mathbf{y} \odot M_{z_1}\mathbf{s}\right) + M_{z_2}\left(\mathbf{y} \odot M_{z_2}\mathbf{s}\right) \right]}_{\mathbf{x}_S} + \\
& + \ \tilde{\gamma}_A \underbrace{\left[M_{z_1}\left(\mathbf{y} \odot M_{z_1}W_{h_A}\mathbf{y}\right) + M_{z_2}\left(\mathbf{y} \odot M_{z_2}W_{h_A}\mathbf{y}\right) \right]}_{\mathbf{x}_A} + \\
& + \ \tilde{\gamma}_R \underbrace{\left[M_{z_1}\left(\mathbf{y} \odot M_{z_1}W_{h_R}\mathbf{y}\right) + M_{z_2}\left(\mathbf{y} \odot M_{z_2}W_{h_R}\mathbf{y}\right) \right]}_{\mathbf{x}_R} + \\
& + \ \tilde{\gamma}_D \underbrace{\left(M_{z_1 z_1} + M_{z_2 z_2}\right)\mathbf{y}}_{\mathbf{x}_D} + \\
& + \ \tilde{\rho}_S M_S \Delta_\tau \mathbf{y} + \tilde{\rho}_A M_A \Delta_\tau \mathbf{y} + \tilde{\rho}_R M_R \Delta_\tau \mathbf{y} + \tilde{\rho}_D M_D \Delta_\tau \mathbf{y} + \\
& + \ \boldsymbol{\epsilon},
\end{aligned}
\tag{26}
$$

where $\boldsymbol{\epsilon}$ is a spatially correlated error. As already said, a feasible way to model $\boldsymbol{\epsilon}$ inspired by the spatial econometric literature is:

$$
\boldsymbol{\epsilon} = \lambda W_\epsilon \boldsymbol{\epsilon} + \boldsymbol{\eta},
\tag{27}
$$

where the matrix W_ϵ should reflect the structure of the spatial dependence of $\boldsymbol{\epsilon}$, while λ its intensity, and, finally, $\boldsymbol{\eta}$ is a vector of well-behaved i.i.d. stochastic components. Unfortunately, the theory is silent on the shape of W_ϵ; therefore, its specification will be data driven and such that it ensures the following properties: i) W_ϵ has zeros on its main diagonal; ii) W_ϵ is symmetric; iii) the spatial correlation is allowed to be nonlinear with respect to the distance (see Appendix B). Numerical simulations in Section 3.5 confirm the presence of spatial dependence in $\boldsymbol{\epsilon}$ and also the goodness of our proposed procedure for the specification of W_ϵ.

3.3   Some remarks on SARD model

The methodology used to estimate the parameters of Eq. (1) strongly departs from the typical approach used in growth empirics, i.e. a log-linear approximation around the long-run equilibrium to obtain a reduced-form model, that is linear with respect to the model parameters (Durlauf et al., 2005). In particular, both the difficulty of characterizing the *spatial equilibrium* and the joint presence of salient *non-linear dynamics* and *spatial spillovers* are the key properties of Eq. (1). Therefore, the SARD econometric Model (4) has been derived with no reference to the equilibrium spatial distribution, but it results linear in the parameters since the discretization over space and time preserves the linearity. Finally, the relationship between the local growth and the initial local level of income $\mathbf{y}_0$ is strongly non linear as the outcome of aggregative, repulsive and diffusive forces.

As regards the space, it can be modelled either in a continuous two-dimensional plane (longitude and latitude) or as a set of discrete spatial units; in the latter case, the *spatial structure* can be represented by connecting arcs, reflecting the interaction between couples of spatial units (Mendes and Mendes, 2015). Spatial econometrics generally adopts the discrete approach to match the only availability of discrete spatial data, and models the spatial structure by the *(spatial) weight matrix*. In particular, the weight matrix identifies if two spatial units are neighbours and, also, the intensity of their interaction; the proximity might be specified in several ways, for example using geographical distance and/or spatial contiguity (LeSage and Pace, 2009).[11] The spatial structure implied by Eq. (1) is, however, more complex since it is expressed in continuous space, which prevents the possibility of specifying the pairwise interactions by a simple weight matrix. Instead, the spatial structure is specified by a function depending on the relative position between two spatial units, and by another function describing the geographical landscape. The latter specifies the intrinsic characteristics of the territory (e.g. the presence of rivers, mountains, etc.), that may affect the strength of the connection between two spatial units. The former, instead, describes how the intensity of the interaction between two units depends on their reciprocal distance, keeping into account also the direction to move from one unit to the other. This effect of directionality is measured by *spatial gradients*, i.e. spatial partial derivatives. The discrepancy between the continuous locations of Eq. (1) and the actual discrete data imposes an intermediate step of discretization to approximate partial derivatives to arrive at the econometric model. In particular, we employ the *generalized finite difference*

---

[11]This choice should be driven by the "problem being modelled, and perhaps particular additional non-sample information which may be available" (Mendes and Mendes, 2015).

method, which is based on a Taylor approximation to compute partial derivatives on irregular grids (Jensen, 1972).[12]

As regards time, continuous time and spatial spillovers induce a contemporaneous effect between two locations. In the discrete time approximation, the final outcome is a model specification which strongly resembles the class of spatial lag models (LeSage and Pace, 2009). This implies that endogeneity is a key feature in the estimate.

Finally, Model (4) resembles typical spatial econometric models, such as the spatial autoregressive model (SAR), the spatial error model (SEM), and also a flavour of an SLX model (LeSage and Pace, 2009). However, the calculation of the matrices Ms, i.e. the counterparts of the spatial weights matrix, is strictly driven from the discretization of Eq. (1) and not a priori assumed. In particular, the model in its parts of spatial lag and error is not derived by strategic decision of rational agents as in Brueckner (2003) or other types of strategic social interaction as in Brock and Durlauf (2001a,b), but as results of the discretization over time and space of a reallocation process affected by exogenous factors (topography), purely random component (diffusion) or the endogenous forces of aggregation and repulsion. This feature is crucial for the identification of the model parameters. As discussed in Gibbons and Overman (2012), the latter is crucially based on the use of the "true" spatial matrix in the estimate, which in our case is driven by the theoretical model and exogenous. Moreover, since Model (4) falls back into the Blume et al. (2015)'s case of "known sociomatrices", its identification seems to be guaranteed from the shape of the matrices Ms.

3.4   Caveats on SARD estimation

The proposed methodology has some challenges and intrinsic limitations discussed below. Firstly, the estimation of Model (4) requires a Maximum Likelihood (ML) or Instrumental Variable (IV) procedure given the presence of endogenous regressors $M_S \Delta_\tau \mathbf{y}$, $M_A \Delta_\tau \mathbf{y}$, $M_R \Delta_\tau \mathbf{y}$ and $M_D \Delta_\tau \mathbf{y}$. On the one hand, ML can be computationally intensive given that one needs to optimize over four parameters (even five in the presence of spatial dependence in the error term) and the normality of the error term is not guaranteed given the presence of nonlinear spatial dependence. In this regard, the simple OLS estimate provides a benchmark robust to non-normality of residuals (see Section 3.5). On the other hand, the IV requires the specification of

---

[12]Appendix E contains a self-contained overview of the method, which amounts to constructing a set of matrices that approximate the derivative operators up to the appropriate order in space.

not weak and valid instruments. Unfortunately, in Section 3.5 we will show that the validity of the usual instruments used in spatial econometrics, based on spatially lagged variables of higher order, is not always guaranteed in our framework.

Secondly, the discretization procedure can introduce a bias in the model, which we investigate in detail in Section 3.5. However, discretization errors are proportional to the distance between the evaluation points and, therefore, bias can be overlooked with a sufficiently fine discretization. In this respect, if the geographical resolution of data is low, i.e. the distance between spatial locations is too wide, then a substantial bias cannot be avoided. As a concrete example, the proposed methodology would most likely fail (or produce meaningless results) if applied to countries' GDP per capita spatial distribution. Also, the proposed framework cannot provide any diagnostic on the minimum resolution needed to get reliable estimates. However, a mismatch between the expected sign of parameters and its estimate could be a signal that the spatial resolution of the data used in the analysis is too low (or that the model is misspecified). In this respect, attention must be given to the fact that parameters are identifiable up to the scale factor $(1 - \tau \rho_\varphi/2)$. Given that $\rho_\varphi \approx \varphi$ and φ is approximately the instantaneous growth rate of y, one should keep into consideration the expected instantaneous growth rate to identify possible changes in the expected signs. The magnitude of the spatial correlation among the units of observation is another important feature to evaluate the bias from discretization. If they are highly correlated, that is neighbouring locations don't differ too much among themselves, then the bias is limited even if locations are spread far apart. On the contrary, the approximation may be very poor even when the units of observations are of small size if the spatial correlation is very low.

Thirdly, the theoretical model also imposes limitations on the possible choice of the (kernel) matrices W_{h_A} and W_{h_R}. The latter derives from the space discretization of the kernel functions K_{h_A} and K_{h_R} in Eq. (1). Accordingly, they require to be defined in terms of *relative distance* between spatial units and, therefore, they are not allowed to be based on contiguity between areas of observations (as, e.g., N-order contiguity).

Fourthly, since the ML estimation of the SARD model is computationally challenging, an educated choice of the initial conditions for the numerical optimization is advised. In this respect, the initial conditions will be set based on an instrumental variable (IV) estimate, which is easy to compute and does not rely on the assumption of normality of the disturbances, although it can be less efficient than the ML estimator (Kelejian and Prucha, 1998). To instrument the spatially

lagged dependent variable, Kelejian et al. (2004) suggests to use $[X \ WX \ldots W^g X]$, where g is usually taken equal to 1 or 2 according to the type of model estimated. In the same vein, we use as instruments $[M_S^2 X \ M_A^2 X \ M_R^2 X \ M_D^2 X]$.

Finally, since the econometric model is obtained by the discretization of a continuous space model, all variables must share the same scale with respect to the unit of observation; for example, the total income of a region must be rescaled by its area.

3.5 Numerical investigations

In this section, we examine the properties of the proposed estimation method of spatial PDEs by numerical investigations. We focus only on the errors induced by discretization and neglect other sources of randomness in ϵ, such as the finite number of individuals and firms and the presence of measurement errors in y.

3.5.1 The baseline setting

As baseline setting, consider a two-dimensional torus of area one, the kernels in Eq. (3) with $h_A = 0.15$ and $h_R = 0.4$, an initial distribution y_0 with three peaks, $\gamma_S = 0$ (no effect from spatial exogenous heterogeneity), $\alpha = 0.01$, $\varphi = 0.01$, $\gamma_A = -0.00175$, $\gamma_R = 0.0025$, and $\gamma_D = 0.00525$. The parameters and the initial distribution are chosen such that the time-space evolution of y simulated by Eq. (1) displays the three spatial behaviours of aggregation, repulsion, and diffusion. For

(a) The initial spatial distribution y_0 in continuos space.

(b) The final spatial distribution $y_{0.1}$ in continuos space.

(c) The spatial distribution of $\Delta_{0.1} y$ in continuos space.

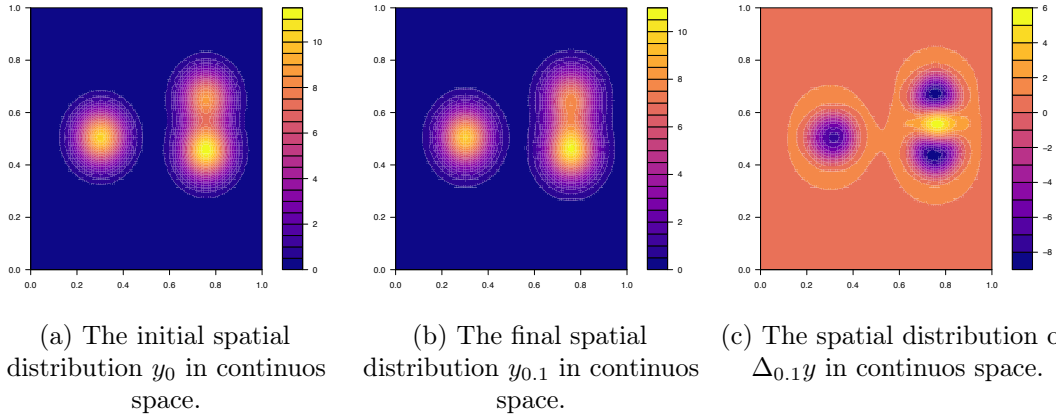Figure 1: Panel (a) reports the initial spatial distribution; Panel (b) the final spatial distribution; and Panel (c) the distribution of time change of y with $\tau = 0.1$ in the 2-dimensional torus of area one. The colour reflects the value of the variable in the location (the blue corresponds to the lowest one).

testing the importance of the time discretization, we consider two-time intervals,

one with $\tau = 0.1$, reported in Figure 1, and one with $\tau = 1$, reported in Figure 2.[13] After 0.1 periods, two of the three peaks on the right in y_0 are merging as the prevalence of the aggregative force; the third peak on the left instead becomes flatter, signalling the prevalence of repulsive and diffusive forces (see Figures 1b and 1c). However, after one period the aggregative force led to a complete merge of the two peaks on the right, while the one on the left spread even more due to repulsive and diffusive forces (see Figure 2b and 2c).

(a) The initial spatial distribution y_0 in continuos space.

(b) The final spatial distribution y_1 in continuos space.

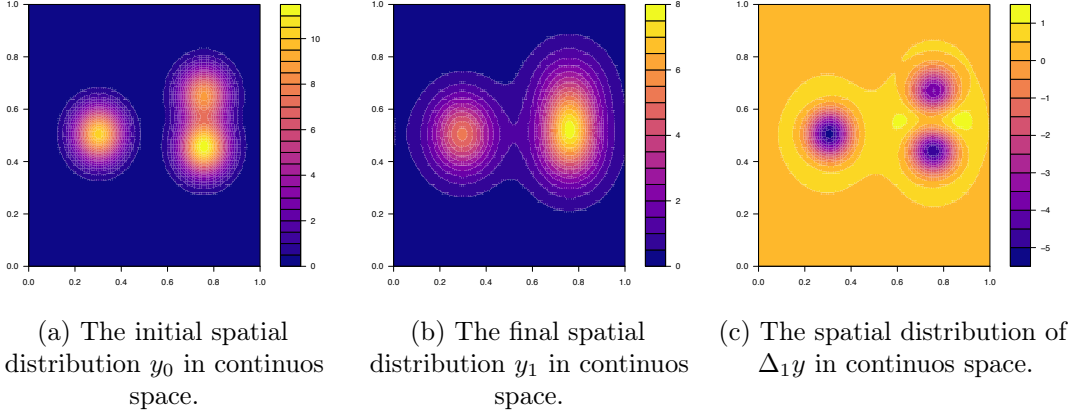(c) The spatial distribution of $\Delta_1 y$ in continuos space.

Figure 2: Panel (a) reports the initial spatial distribution; Panel (b) the final spatial distribution; and Panel (c) the distribution of time change of y with $\tau = 1$ in the 2-dimensional torus of area one. The colour reflects the value of the variable in the location (the blue corresponds to the lowest one).

To appreciate the impact of the space discretization, Figures 3a, 3b and 3c depict the case of a partition of space into $N = 144$ regular cells for $\tau = 1$. Given this discretization, Figures 3d, 3e and 3f report the maps of regressors x_A, x_R and x_D respectively. Given that $\gamma_A < 0$ while $\gamma_R > 0$, the attractive and repulsive forces show a similar spatial pattern but an opposite effect on the dynamics of y: the first force induces the emergence of one peak on the right, while the second works to flatten the peak on the left. Finally, the diffusive force has a particularly strong (negative) magnitude in correspondence of three peaks.

3.5.2 The parameters' estimate

To test the capacity of our methodology to identify the parameters of Eq. (1) under different geographic granularity of data, consider five discretizations of space, i.e $N \in \{144, 400, 900, 1600, 2500\}$ and five possible time intervals, i.e. $\tau \in \{0.1, 0.25, 0.5, 0.75, 1\}$. Figure 4 is a graphical illustration of all the combinations

---

[13]With a little abuse of notation we will denote by $\tau = 0.1$ the case where $\tau = 1$ and $\Delta = 0.1$.

(a) The initial spatial distribution y_0 partitioned into 144 cells.

(b) The final spatial distribution y_1 partitioned into 144 cells.

(c) The spatial distribution of $\Delta_1 y$ partitioned into 144 cells.

(d) The spatial distribution of x_A partitioned into 144 cells.

(e) The spatial distribution of x_R partitioned into 144 cells.

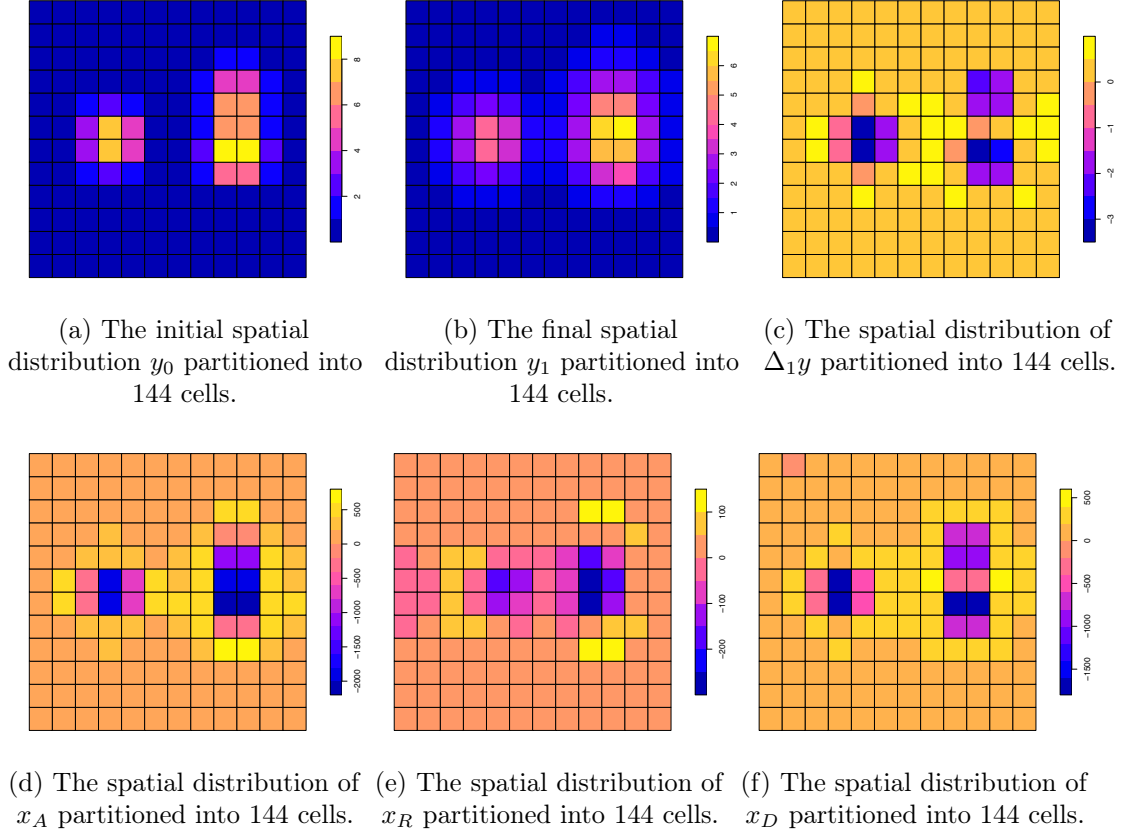(f) The spatial distribution of x_D partitioned into 144 cells.

Figure 3: For a partition of space into $N = 144$ regular cells, Panel (a) reports the initial spatial distribution; Panel (b) the final spatial distribution; Panel (c) the distribution of time change of y with $\tau = 1$; Panel (d) the spatial distribution of x_A; Panel (e) the spatial distribution of x_R; and Panel (f) the spatial distribution of x_D of Model (4). The colour reflects the value of the variable in the units of observation (the blue corresponds to the lowest one).
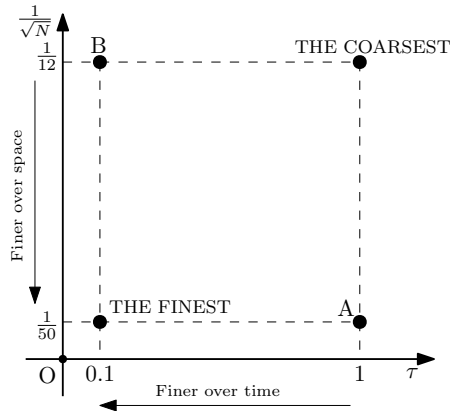

Figure 4: The possible combinations in the discretization over time and space when $N \in \{144, 400, 900, 1600, 2500\}$ and $\tau \in \{0.1, 0.25, 0.5, 0.75, 1\}$.

of discretization over time and space, where the origin of the axes O represents the case of continuous time and space, the combination $N = 2500$ and $\tau = 0.1$ the finest discretization, while the one with $N = 144$ and $\tau = 1$ the coarsest.

We consider four types of estimate: i) OLS NAIVE, i.e. the ordinary least squares estimate of Model (4) without any control for the bias due to discretization (i.e. imposing $\rho_S = \rho_A = \rho_R = \rho_D = 0$); ii) OLS, i.e. the ordinary least squares estimate of Model (4), which means including the controls for the bias due to discretization without dealing with their endogeneity; iii) IV, i.e. the instrumental variables estimate of Model (4), which means including the controls for the bias due to discretization and dealing with their endogeneity, using as instruments the second order spatially lagged exogenous regressors; and finally, iv) ML, i.e. the maximum likelihood estimate of Model (4), taking errors specified as in Eq. (10), which means including the controls for the bias due to discretization and dealing with their endogeneity and, in addition, accounting for the presence of spatial correlation in the residuals.

The comparison between the finest and the coarsest discretization models reported in Table 1 highlights the improvement in the estimated parameters both in terms of bias and overall fitness as measured by the AICc and Mean Squared Error (MSE). Across the four estimation methods, ML outperforms all other methods in terms of AICc although it has a modestly higher bias with respect to OLS and almost no difference in the MSE. Moreover, in the ML estimation, the true values are always within the 95% confidence intervals. The use of IV results in a poor estimation both in terms of AICc, MSE and bias. The closeness of the performance of ML and OLS highlights that for our setting, the endogeneity in the time discretization controls as well as the possible departure from the normality of the errors are not too strong. While the ML estimated parameters are always statistically significant at the usual levels, the OLS and IV are no longer significant when we estimate the standard errors via non-parametric bootstrap (i.e. no assumptions on the error component).

Finally, Figure 5 confirms that ML estimation outperforms all three other types of estimation in terms of AICc for alternative discretizations starting from the coarsest case of Figure 4 (see Appendix F for the point estimate of the model parameters). Overall, OLS and IV appear to be a valid alternative to ML in some circumstances.

	Parameters	$\alpha = 0.01$	$\varphi = 0.01$	$\gamma_A = -0.00175$	$\gamma_R = 0.0025$	$\gamma_D = 0.00525$	AICc	MSE
N	τ				**OLS NAIVE**				
144	1	Mean	0.16797	-0.14647	-0.00097	0.00054	0.00327	-228.7	0.01096
		Bias	0.15797	-0.15647	0.00078	-0.00196	-0.00198		
		SE	0.01185	0.00784	0.00005	0.00023	0.00006		
		SE^B	0.69501	0.58907	0.00437	0.00279	0.01396		
		RMSE	0.69265	0.59059	0.00863	0.00681	0.01419		
2500	0.10	Mean	0.03659	-0.01804	-0.00136	0.00175	0.00426	-6290.2	0.00471
		Bias	0.02659	-0.02804	0.00039	-0.00075	-0.00099		
		SE	0.00168	0.00097	0.0000	0.00002	0.00001		
		SE^B	0.05900	0.02909	0.00222	0.00285	0.00693		
		RMSE	0.06700	0.03747	0.00825	0.00637	0.00836		
					OLS				
144	1	Mean	0.04803	-0.03236	-0.00119	0.00139	0.00406	-556.1	0.00589
		Bias	0.03803	-0.04236	0.00056	-0.00111	-0.00119		
		SE	0.01467	0.01311	0.00019	0.00054	0.00036		
		SE^B	0.19821	0.14384	0.01107	0.01083	0.02988		
		RMSE	0.20220	0.14835	0.01359	0.01225	0.03007		
2500	0.10	Mean	0.00855	0.01111	-0.00175	0.00246	0.00521	-19023.4	0.00025
		Bias	-0.00145	0.00111	0.00000	-0.00004	-0.00004		
		SE	0.00044	0.00031	0.00000	0.00001	0.00000		
		SE^B	0.01347	0.01727	0.00273	0.00383	0.00812		
		RMSE	0.01464	0.01884	0.00870	0.00656	0.00934		
					IV				
144	1	Mean	0.29173	-0.26550	-0.00247	-0.00167	0.00643	-408.9	0.01740
		Bias	0.28173	-0.27550	-0.00072	-0.00417	0.00118		
		SE	0.06446	0.06346	0.00057	0.00173	0.00131		
		SE^B	0.73241	0.65902	0.01480	0.01617	0.03611		
		RMSE	0.75834	0.68652	0.01664	0.01731	0.03510		
2500	0.10	Mean	0.01757	0.00169	-0.00174	0.00245	0.00510	-18985.7	0.00032
		Bias	0.00757	-0.00831	0.00001	-0.00005	-0.00015		
		SE	0.00179	0.00176	0.00000	0.00001	0.00001		
		SE^B	0.03353	0.02036	0.00275	0.00388	0.00808		
		RMSE	0.03613	0.02101	0.00870	0.00659	0.00930		
					ML				
144	1	Mean	-0.06940	0.08946	-0.00392	0.00600	0.00916	-661.8	0.01056
		Bias	-0.07940	0.07946	-0.00217	0.00350	0.00391		
		SE	0.00684	0.00677	0.00007	0.00025	0.00016		
		RMSE	0.07178	0.08051	0.01003	0.00454	0.00575		
2500	0.10	Mean	0.00740	0.01256	-0.00179	0.00252	0.00532	-20611.7	0.00026
		Bias	-0.00260	0.00256	-0.00004	0.00002	0.00007		
		SE	0.00115	0.00115	0.00015	0.00021	0.00045		
		RMSE	0.00516	0.00874	0.00833	0.00524	0.00453		

Table 1: The estimate of Model (4) parameters for the discretization $N = 2500$ and $\tau = 0.1$ (the finest) and $N = 144$ and $\tau = 1$ (the coarsest), using four estimation methods: OLS NAIVE (ordinary least squares without controls for time discretization); OLS (ordinary least squares); IV (instrumental variables); and ML (maximum likelihood). SE are standard errors; SE^B non-parametric bootstrap standard errors; $RMSE$ root mean squared error; AICc (corrected) Akaike information criterion; MSE: Mean Squared Error.
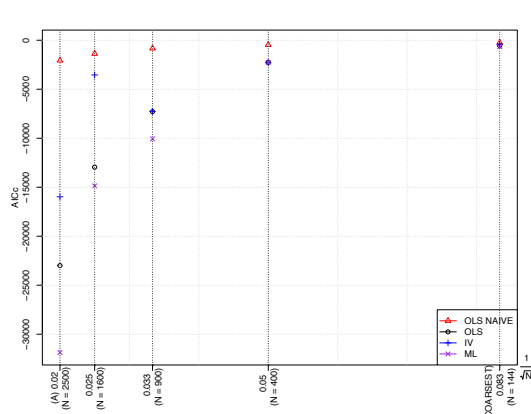
(a) AICc for alternative spatial discretization and type of estimation starting from the coarsest case of Figure 4 and increasing the finesse of spatial discretization in the case $\tau = 1$.
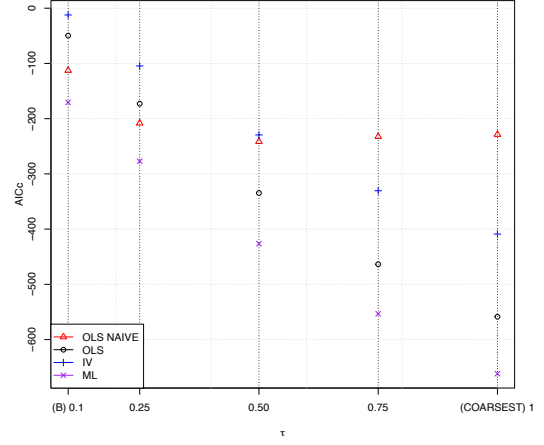
(b) AICc for alternative time discretization and type of estimation starting from the coarsest case of Figure 4 and increasing the finesse of time discretization in the case $N = 144$.

Figure 5: AICc for different levels of time and spatial discretization for the four estimation methods: OLS NAIVE (ordinary least squares without any control for time discretization); OLS (ordinary least squares); IV (instrumental variables); and ML (maximum likelihood).

N	τ	$\hat{\lambda}$	$\hat{\ell}_1$	$\hat{\ell}_2$	$\hat{\ell}_3$	$\hat{\ell}_4$	$\hat{\ell}_5$	$\hat{\ell}_6$	$\hat{\ell}_7$	$\hat{\ell}_8$	$\hat{\ell}_9$	$\hat{\ell}_{10}$
144	1	0.622	-2.743	-2.494	-2.151	-1.861	-1.537	-1.239	-0.973	-0.691	-0.477	-0.190
		(0.068)	(0.380)	(0.335)	(0.297)	(0.256)	(0.217)	(0.179)	(0.142)	(0.109)	(0.078)	(0.041)
2500	0.10	0.775	0.159	-0.049	0.024	-0.007	0.011	-0.002	0.003	0.004	-0.001	0.002
		(0.011)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)

Table 2: The estimate of λ and ℓs used for the calculation of W_ϵ in the ML estimate following the procedure described in Appendix B; $\hat{\ell}_i - \hat{\ell}_{i+1}$ measures the importance of spatial dependence among cells with order of contiguity i. Standard errors are reported between brackets.

3.5.3 The spatial dependence in the reminder

The significant and positive value of λ in the ML estimation reported in Table 2 shows evidence in favour of the presence of spatial dependence in the error term. The same table reports the estimated ℓs used in the estimate of W_ϵ with the procedure described in Appendix B. While the estimate for the coarsest discretization is not satisfying, for the finest we have the expected outcome that the first order of contiguity is dominant with respect to the other orders, with a maximum significant order equal to 5 and an estimated λ positive and close to 1 (see Appendix B).

(a) The spatial distribution of the reminder from the theoretical model

(b) The spatial distribution of the remainder after the filtering by W_ϵ.

(c) The estimated spatial correlogram (Moran's I) of the reminder for alternative orders of spatial contiguities (with their 95% confidence bands).
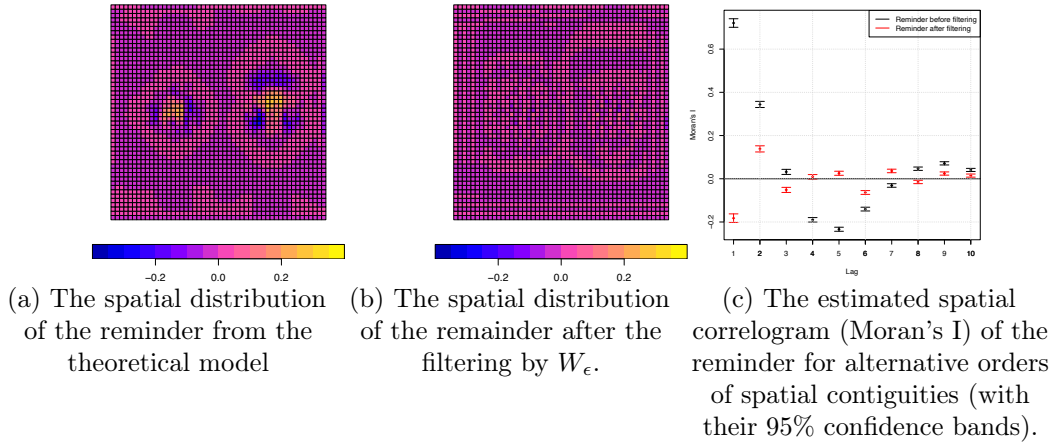
Figure 6: The reminder in the finest case ($N = 2500, \tau = 0.1$).

Figure 6a reports the spatial distribution of the reminder in the finest case ($N = 2500, \tau = 0.1$) directly calculated from the theoretical model assuming that $\rho_j = (\tau/2)\,\gamma_j$ for $j \in \{A, R, D\}$. As expected, the stronger spatial dependence (both positive and negative) is present around the three peaks. Figure 6b, instead, reports the spatial distribution of filtered reminders, where filtering is made by W_ϵ estimated applying on the reminders the procedure described in Appendix B taking $\lambda = 1$. A comparison between Figures 6a and 6b confirms the ability of the proposed procedure to remove the most of spatial dependence. Such ability is further confirmed by the estimate of the spatial correlogram (Moran's I) of the reminder (before and after the filtering) for different orders of spatial contiguities reported in Figure 6c.

4 Empirical application

In this section, we estimate Model (4) for personal income of Italian municipalities in the period 2008-2019 and compare our estimate with the most common spatial econometric models.

4.1 Data on Italian municipalities

Italian Ministry of Economy and Finance (Agenzia delle Entrate) releases information at the municipal level (about 8000 in Italy) on the *nominal personal income* declared for tax purposes (IRPEF) for the period 2008-2020 from resident households.[14] During this period, Italy was hit by three main shocks: the subprime mortgage crisis coming from the US in 2007-2011, the sovereign debt crisis started in 2011-2013, and the COVID pandemic in 2020. Given the peculiarities of the COVID pandemic, we exclude 2020. In the analysis, personal income is not corrected for inflation because the latter is not available at the municipal level. However, in the period 2008-2019, the inflation was extremely low, with an annual average inflation rate at the national level of about 1.2%, and with marginal regional heterogeneity, with the most of Italian region having a rate in the range [1%,1.7%].[15]
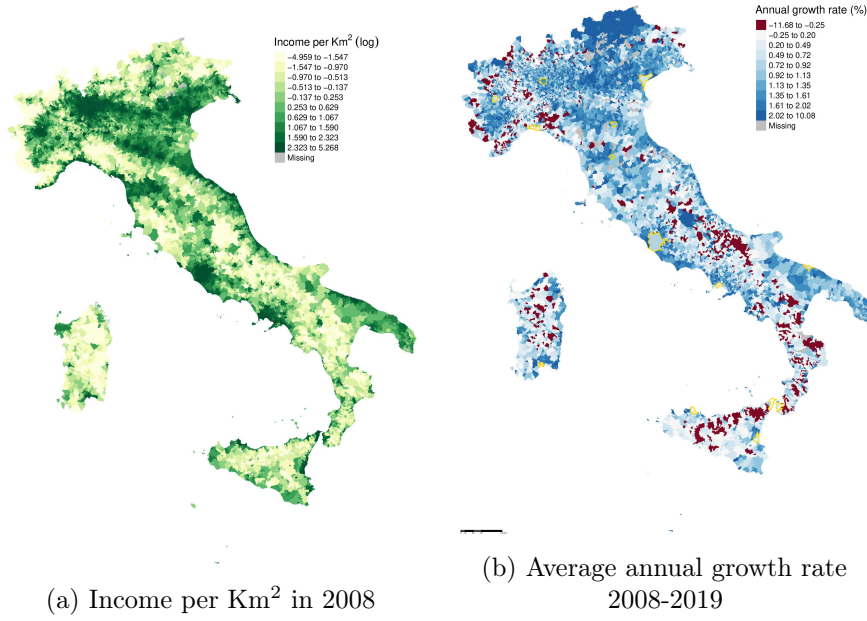
To make comparable the income of municipalities with very different sizes and different types of territory, the total personal incomes are divided by the area of each municipality, measured in Km^2. Figure 7 reports the map of total income per Km^2 of 7807 Italian municipalities for the year 2008 and its annual growth rate over the period 2008-2019. The general impression is of a high level of agglomeration in the spatial income distribution, which seems to be persistent over time. There also exists a remarkable difference between the Northern and Southern parts of Italy, with the highest level of agglomeration in the North, between the inner and coastal areas and finally, the key impact of mountains which hurts the agglomeration process. In this respect, the estimate will include also the (average) altimetry of municipalities, gathered from ISTAT.

4.2 The estimate of SARD model

In the empirical application, we consider $\tau = 11$ and several alternative values both for h_A and h_R of Eq. (3). In particular, we probe all the combinations of

---

[14]https://www1.finanze.gov.it/finanze/pagina_dichiarazioni/public/dichiarazioni.php

[15]Source: ISTAT (*Italian National Institute of Statistics*).

Income per Km² (log)
- −4.959 to −1.547
- −1.547 to −0.970
- −0.970 to −0.513
- −0.513 to −0.137
- −0.137 to 0.253
- 0.253 to 0.629
- 0.629 to 1.067
- 1.067 to 1.590
- 1.590 to 2.323
- 2.323 to 5.268
- Missing

Annual growth rate (%)
- −11.68 to −0.25
- −0.25 to 0.20
- 0.20 to 0.49
- 0.49 to 0.72
- 0.72 to 0.92
- 0.92 to 1.13
- 1.13 to 1.35
- 1.35 to 1.61
- 1.61 to 2.02
- 2.02 to 10.08
- Missing

(a) Income per Km² in 2008

(b) Average annual growth rate 2008-2019

Source: Italian Ministry of Economy and Finance (Agenzia delle Entrate) and ISTAT (*Italian National Institute of Statistics*).

Figure 7: The map of the (log of) total personal income per Km² of Italian municipalities in 2008 and annual growth rates over the period 2008-2019. Yellow borders indicate the 14 Italian metropolitan municipalities.

$h_A \in \{5, 10, 15, 20\}$ and $h_R \in \{25, 30, \ldots, 95, 100\}$ (Km), which overall amounts at estimating 64 models. Table 3 only reports the result of the best combination, that is the combination with the lowest AICc.

SARD model in the first column of Table 3 is able to explain almost 77% of the cross-sectional variance of the income change of Italian municipalities over the period 2008-2019. The coefficient $\widehat{\widetilde{\varphi}}$, measuring the impact of the initial level of income on its variation, is significant and equal to 0.013. From the comparison of $\widehat{\widetilde{\varphi}}$ and the long-run aggregate growth rate calculated by Eq. (2), is it possible to calculate the value of $\widehat{\rho}_\varphi$ and, consequently, the scale factor (equal to 0.8282) needed to retrieve the parameters of Eq. (1) from the estimated coefficients of Model (4).[16]

The estimated coefficient of topography $\widehat{\widetilde{\gamma}}_S$ is positive and significant, i.e. a nonuniform topography hurts economic activity. This finding is in line with the well-known phenomenon of depopulation characterising the mountain areas of Italy (ISTAT, 2017). Although the estimated $\widehat{\widetilde{\gamma}}_S$ appears very low in absolute value, the analysis in Section 4.4 shows its substantial impact on the spatial distribution of

---

[16]For example, $\widehat{\gamma}_S$ is equal to 1.08e-05 × 0.8282 = 8.94e-06.

Italian income density.

The estimated coefficients of aggregation and repulsion ($\widehat{\widetilde{\gamma}}_A$ and $\widehat{\widetilde{\gamma}}_R$, respectively) display the signs suggested by Eq. (1), and are both statistically significant at the usual levels. In absolute value, $\widehat{\widetilde{\gamma}}_A$ is marginally but significantly higher than $\widehat{\widetilde{\gamma}}_R$. The aggregation is therefore prevailing on repulsion although aggregation forces work within a smaller radius given that the best specification for the spatial structure is given by $h_A = 10$ km and $h_R = 45$ km. Overall, this implies the emergence of urban agglomerates, but whose radius should be limited within 10 km (Rosenthal and Strange, 2008). At the same time, the crowding/congestion effects, working in a radius of 45 km, explain the extension of suburban areas and commuting zones (Alonso, 1964; Duranton and Puga, 2020; Glaeser et al., 2001; Kneebone, 2009; Travisi et al., 2010).

The estimated coefficient $\widehat{\widetilde{\gamma}}_D$ has the expected positive and significant sign, highlighting the presence of diffusive forces in the spatial distribution of income density which tend to equalise it. A detailed quantitative assessment of these forces is left to Section 4.4. The coefficients for the corrections of time discretization are all highly significant.[17]

Finally, the spatial dependence in the error, as measured by the coefficient $\hat{\lambda}$, is positive and significant, and is close to one. Given that all elements of W_ϵ are non-negative (see Appendix G), in our specific application the error exhibits positive spatial correlation. This means that neighbouring municipalities' dynamics are still positively correlated even after controlling for topographical, aggregative, repulsive, and diffusive forces. This positive correlation could also result from a (unaccounted) place-based policy that affects multiple neighbouring municipalities, leading agents (individuals/firms) to coordinate their spatial movements.

The SARD model lacking spatial dependence in the errors exhibits inferior performance in terms of AICc and R^2-N, as demonstrated by the WN-SARD model estimated through either maximum likelihood (ML) or instrumental variables (IV). The estimated IV coefficients have marginally different magnitudes, yet they retain their respective signs and levels of significance. Instead, omitting the corrections for time discretization (NAIVE SARD) results in a significant decrease in model performance and discrepancies in the estimated coefficients, despite retaining their signs and significance.

---

[17]In particular, the estimated five coefficients are $\widehat{\widetilde{\rho}}_S = -8.27\text{e-}04$, $\widehat{\widetilde{\rho}}_A = 1.99\text{e-}06$, $\widehat{\widetilde{\rho}}_R = -1.96\text{e-}06$ and $\widehat{\widetilde{\rho}}_D = -5.97\text{e-}05$.

Dependent variable: $\Delta_{11}Y$

	SARD (ML)	WN SARD (ML)	WN SARD (IV)	WN SARD (OLS)	NAIVE SARD (OLS)	INCOME LAG (OLS)	ALT INCOME LAG (OLS)	S INCOME LAG (OLS)	SLX (OLS)	SPATIAL LAG (ML)	SPATIAL DURBIN (ML)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
$\tilde{\alpha}$	1e-03** (6.87e-04)	5.91e-04** (3.18e-04)	1.27e-03* (6.64e-04)	1.01e-03* (5.93e-04)	1.66e-03*** (7.85e-04)	6.31e-03*** (8e-04)	0.012*** (1.21e-03)	6.41e-03*** (7.94e-04)	-3.33e-03*** (1.37e-03)	-3.08e-03*** (1.14e-03)	-2.87e-03** (1.28e-03)
$\tilde{\varphi}$	0.013*** (5.98e-05)	0.013*** (4.14e-05)	0.013*** (9.13e-05)	0.013*** (7.79e-05)	0.013*** (1.02e-04)	0.011*** (8.76e-05)	0.011*** (9e-05)	0.011*** (8.75e-05)	7.58e-03*** (1.45e-04)	7.48e-03*** (1.36e-04)	7.90e-03*** (1.36e-04)
γ_{ALT}							-0.011*** (1.74e-03)		-0.017*** (4.95e-03)	-2.49e-03 (1.56e-03)	-0.015*** (4.62e-03)
θ									6.32e-03*** (2.09e-04)		-2.24e-03*** (3.36e-04)
θ_{ALT}									0.015 (5.93e-03)		0.015 (5.54e-03)
$\tilde{\gamma}_S$	1.08e-05*** (2.17e-07)	1.08e-05*** (1.98e-07)	8.88e-06*** (6.65e-07)	1.15e-05*** (4.61e-07)	7.67e-06*** (4.88e-07)			5.4e-06*** (4.89e-07)			
$\tilde{\gamma}_A$	-3.55e-01*** (2.93e-03)	-4.3e-01*** (2.99e-03)	-3.4e-01*** (0.017)	-4.81e-01*** (0.012)	-1.7e-01*** (7.39e-03)						
$\tilde{\gamma}_R$	3.397*** (0.049)	3.043*** (0.044)	3.616*** (0.529)	4.588*** (0.21)	0.751*** (0.109)						
$\tilde{\gamma}_D$	2.38e-06*** (3.61e-08)	2.19e-06*** (3.69e-08)	2.48e-06*** (1.36e-07)	3.7e-06*** (7.45e-08)	1.61e-06*** (9.11e-08)						
ρ										0.494*** (0.011)	0.607*** (0.019)
λ	0.682*** (0.018)										
Correction for time discretization	YES	YES	YES	YES	NO						
AICc	-23220.5	-22169.6	-21881.4	-21436.3	-21434.3	-20678.3	-20716.8	-20797.5	-21582.8	-22318.4	-22370.4
MSE	0.0022	0.0029	0.0026	0.0021	0.0037	0.0041	0.0041	0.0040	0.0036	0.0032	0.0032
R^2-N	0.7694	0.7360	0.7261	0.7096	0.7096	0.6802	0.6818	0.6851	0.7153	0.7410	0.7427
D	10	10	10	10	10				25	25	25
h_A											
h_R	45	45	45	45	45						

Note: AICc: corrected Akaike Information Criteria. MSE: Mean Squared Error. R^2-N: Nagelkerke Pseudo-R^2. Significance levels: *p<0.1; **p<0.05; ***p<0.01

Table 3: The estimated coefficients of different econometric models for Italian municipality incomes per km^2 for the period 2008-2019: ML estimation of SARD model (1); ML estimation of SARD model without spatially correlated errors (2); IV estimation of SARD model without spatially correlated errors (3); OLS estimation of SARD model without spatially correlated errors (4); OLS estimation of SARD model without spatially correlated errors and corrections for time discretization (5); estimation of a benchmark model with only lagged income (6); estimation of a benchmark model also including altimetry (7); estimation of a benchmark model also including our proxy for topography (8); estimation of SLX model with lagged income and altimetry (9); estimation of SPATIAL LAG model with lagged income and altimetry (10); estimation of SPATIAL DURBIN model with lagged income and altimetry (11). Source: our estimations on data from the Italian Ministry of Economy and Finance (Agenzia delle Entrate).

4.3 A comparison with other (spatial) econometric models

The comparison between the SARD model and the simpler model that solely encompasses the starting income level (INCOME LAG in Table 3, absolute convergence in growth rate) indicates that the $\widehat{\varphi}$ coefficient is consistently stable across both models, but the explained variance reduces to approximately 0.68 in the latter. The incorporation of altimetry in the ALT INCOME LAG model (convergence conditioned to altimetry), despite being negative and significant, does not significantly improve the fitness of the model. A notable improvement is instead reached in the S INCOME LAG model (convergence conditioned to $\mathbf{x}_S$) where we consider a more complex measure of topography such as the $\mathbf{x}_S$ previously employed in the SARD model. However, the fitness of S INCOME LAG model still lags behind that of the SARD model.

Finally, to compare the performance of Model (4) with respect to models typically used in spatial econometrics, we estimate SLX, SPATIAL LAG and SPATIAL DURBIN models, which consider both the starting income level and altimetry and use a spatial weight matrix W whose elements are defined as:

$$w_{ij} = \begin{cases} \frac{1}{d_{ij}^2} & \text{if } d_{ij} \leq \bar{d} \, \forall i \neq j \\ 0 & \text{otherwise,} \end{cases} \qquad (28)$$

where $\bar{d}$ is chosen in the set $\{5, 10, \ldots, 95, 100\}$ kilometres on the base of the lowest AICc. Spatial dependence is present in all the models, with the estimated coefficient for the spatially lagged municipal income change $\hat{\rho}$ being positive and significant (ranging from 0.49 to 0.61), as well as the coefficient of the spatially lagged municipal income $\hat{\theta}$ (ranging from -2.24e-03 to 6.32e-03). The SPATIAL DURBIN model exhibits the best fit, although it remains lower than the SARD model.

4.4 Decomposing the local growth

The impact on the spatial dynamic of the Italian municipal income of each component of the SARD model in Table 3 is calculated by a counterfactual methodology. In particular, for each reallocative component of the SARD model, we calculate a counterfactual in sample average annual growth rate of municipality setting to zero the coefficient related to the component of interest, i.e. $\widehat{\mathbf{g}}_j^{CF} = [(\widehat{\mathbf{y}}^{2019} | \gamma_j = 0) / \mathbf{y}^{2008}]^{1/11} - 1$, where $\widehat{\mathbf{y}}^{2019} | \gamma_j = 0$ is the vector of forecasted municipal incomes in 2019 setting $\gamma_j = 0$, with $j \in \{S, A, R; D\}$. The contribution

of component j is therefore measured by $\widehat{\mathbf{g}}_j = \widehat{\mathbf{g}} - \widehat{\mathbf{g}}_j^{CF}$.

(a) Contribution of topography S

(b) Contribution of aggregation A

(c) Contribution of repulsion R
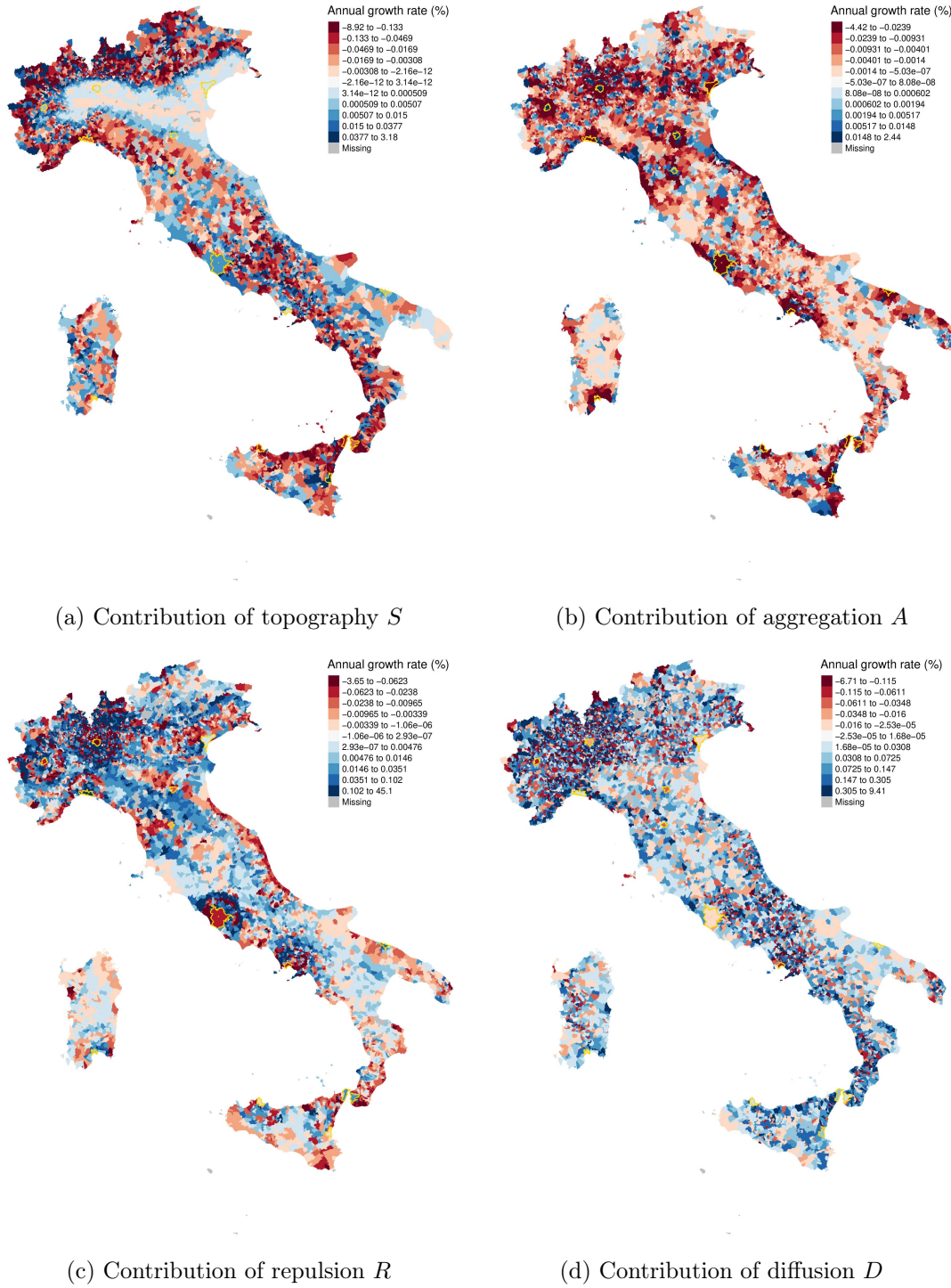
(d) Contribution of diffusion D

Figure 8: The individual contribution of components $\{S, A, R, D\}$ to the forecasted in sample average annual growth rate 2008-2019 based on the estimate of SARD model reported in Table 3. Yellow borders indicate the 14 Italian metropolitan municipalities.

Figure 8a shows that the areas at the bottom of mountains have benefited at most by the topography S, with mountain areas having been mainly hurt, and all flat areas having no effect. Topography appears a driving force for the emergence of medium cities on the borders of the Po Valley and in the central part of the Adriatic coast. Large alpine valleys, such as the ones in Valle d'Aosta, Val di Susa, Trento and Bolzano, also display a positive effect. Finally, the estimated benefit for the municipalities just around Volcanos Vesuvio and Etna's areas is further evidence of the importance of topography. The effect of the aggregation component is spatially heterogeneous, with a prevalent positive contribution for municipalities in coastal areas and the Po Valley (see Figure 8b). However, being a metropolitan municipality does not imply gaining from agglomeration forces because, in the final outcome, a key role is played by the income of neighbouring municipalities acting as a competitive force: Milan and Naples, where the second has very rich neighbouring municipalities with respect to the first, have a gain of 0.31% and a loss of 0.25% from the aggregation component, respectively (see column $W_{10}\mathbf{y}^{2008}$ versus column $\mathbf{y}^{2008}$ and column $\hat{\mathbf{g}}_A$ in Table 4). The extension of Rome cannot allow for a precise identification of an aggregation effect with respect to its centre. The effect of the repulsion component is instead negative for the six richest Metropolitan cities and, in general for the richest municipalities (see Figure 8c and column $\hat{\mathbf{g}}_R$ in Table 4). Finally, the effect of the diffusion component appears random across Italian municipalities (Figure 8d), but negative for the greatest Metropolitan municipalities (Milan, Turin, Naples, Florence, Bologna, Rome, see column $\hat{\mathbf{g}}_D$ in Table 4). Overall, aggregation and repulsion components seem to play the major role for most of the Metropolitan municipalities (Table 4).

4.5 Convergence in local income

Our model allows to account for the sources of (di)convergence in local income. Figure 9 reports the nonparametric estimate (kernel regression) of the relationship between each one of the four components of local growth and the initial level of local income. A positive slope points to that component as a source of divergence, for example component S (topography) represented by the red line in Figure 9 decreases up to a 0.4% in annual growth rate for municipalities with a very low income density in 2008 (e.g. the mountain areas). On the contrary, components R (repulsion) and D (diffusion) (blue and violet lines, respectively) display a negative slope, i.e. they are sources of convergence. In particular, diffusion exerts such convergent impact on the whole range of income, while repulsion only benefits

Metropolitan city	$\mathbf{y}^{2008}$	$W_{10}\mathbf{y}^{2008}$	$W_{45}\mathbf{y}^{2008}$	$\dfrac{W_{10}\mathbf{y}^{2008}}{\mathbf{y}^{2008}}$	$\dfrac{W_{45}\mathbf{y}^{2008}}{\mathbf{y}^{2008}}$	$\hat{\mathbf{g}}$	$\hat{\mathbf{g}}_S$	$\hat{\mathbf{g}}_A$	$\hat{\mathbf{g}}_R$	$\hat{\mathbf{g}}_D$
Milan	164	1006	9502	6.13	57.87	1.22	0.00	0.31	-0.12	-0.02
Turin	115	236	1259	2.06	10.98	1.04	0.01	0.08	-0.03	-0.07
Naples	85	807	3036	9.51	35.76	0.7	0.00	-0.25	-0.04	-0.09
Florence	65	46	392	0.71	6.04	1.01	0.00	0.01	-0.01	-0.06
Bologna	53	76	398	1.41	7.4	1.02	0.00	0.02	-0.02	-0.05
Palermo	46	32	175	0.7	3.83	0.96	-0.06	-0.05	-0.01	0.00
Genoa	43	10	446	0.24	10.33	0.88	-0.03	-0.34	0.16	0.02
Rome	37	0	459	0	12.33	0.98	0.00	-0.03	-0.05	-0.02
Bari	37	52	154	1.41	4.18	0.98	0.00	-0.16	0.05	0.02
Cagliari	29	58	117	1.98	3.97	1.03	0.00	-0.07	0.01	0.02
Catania	17	11	458	0.67	27.05	1.13	0.08	-0.06	-0.01	0.03
Messina	13	10	218	0.75	17.08	0.98	-0.08	-0.07	0.04	0.04
Venice	11	9	902	0.83	84.18	1.01	0.00	-0.06	0.01	0.00
Reggio di Calabria	8	2	220	0.24	26.16	1.04	-0.04	-0.02	0.01	0.00

Table 4: Income per Km² in 2008 ($\mathbf{y}^{2008}$), cumulative income per Km² of municipalities within 10 Kms in 2008 ($W_{10}\mathbf{y}^{2008}$), cumulative income per Km² of municipalities within 45 Kms in 2008 ($W_{45}\mathbf{y}^{2008}$), the fitted growth rate of income per Km² (in %) in the period 2008-2019 ($\hat{\mathbf{g}}$), the counterfactual growth rate of income per Km² (in %) in the period 2008-2019 due to component j, with $j \in \{S, A, R, D\}$ ($\hat{\mathbf{g}}_j$) for the fourteen Italian Metropolitan cities.

low-income municipalities. Finally, component A (aggregation) (green line) is a source of convergence only for very high-income municipalities, i.e. for an income per Km² in 2008 greater than $\exp(3) \approx 20$ (current thousands of euros per squared kilometre), in agreement with the figures in Table 4, where the most of metropolitan municipalities showed a negative contribution from the aggregative component.

4.6 Forecasting the local income

A potential application of the SARD model is forecasting local income. In Figure 10b we map the forecasted income per Km² in 2069 using the SARD model, taking 2019 as the starting year. In the forecast, most of the municipalities are anticipated to grow, with a median growth rate amounting to 1.17% (municipalities with a forecasted negative growth are only 74). A graphical comparison of Figures 10a and 10b suggests that (relative) persistence is the dominant feature for spatial distribution dynamics: in other words, the spatial pattern of income density is already consolidated in Italy, which implies that the spatial distribution in 2019 is close to its long-run equilibrium.
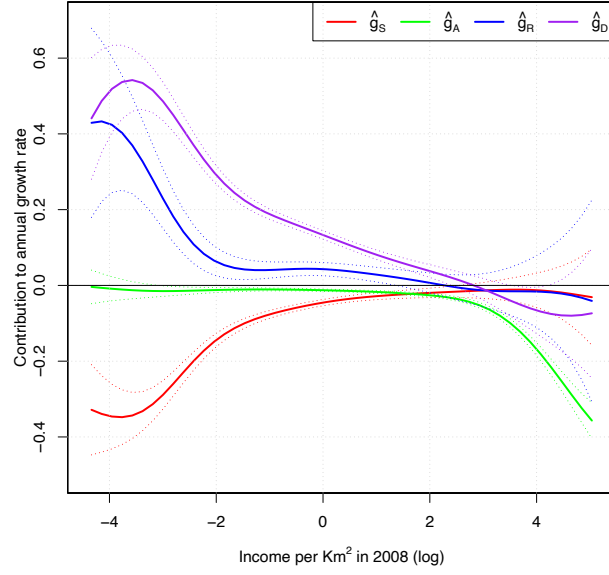
Figure 9: The nonparametric estimate (kernel regression) of the relation between each reallocative component (S, A, R, D) of local growth and the initial level of income (dotted lines represent 95% confidence bands).
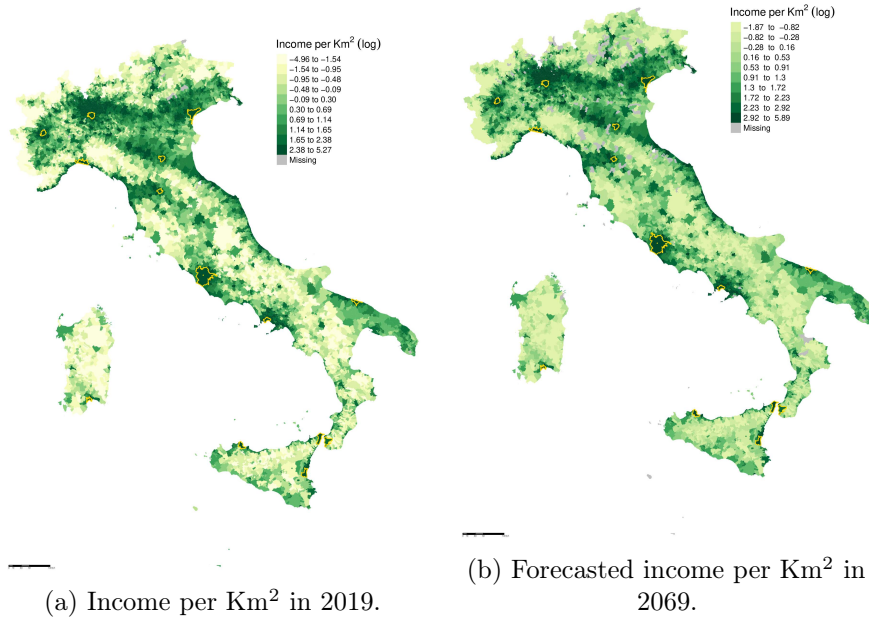

(a) Income per Km^2 in 2019.

(b) Forecasted income per Km^2 in 2069.

Figure 10: The comparison between the income per Km^2 of Italian municipalities in 2019 versus the forecasted income per Km^2 in 2069 based on the estimated SARD model in Table 3. Yellow borders indicate the 14 Italian metropolitan municipalities.

5 Concluding remarks

We have proposed a novel family of spatial econometric models, derived from continuous time-space models defined by the formalism of PDEs. This new family offers an alternative to traditional spatial econometrics (LeSage and Pace, 2009), and the burgeoning field of quantitative spatial economics (Redding and Rossi-Hansberg, 2017).

Our models allow to disentangle the mechanism of *accumulation* from the one of *reallocation*, and enable the identification of four different classes of spatial effects, i.e. *topography*, *aggregation*, *repulsion* and *diffusion*. These effects are identified through spatial matrices derived from our theoretical framework and computed based on the spatial distribution of observations. Consequently, our methodology overcomes the traditional identification issues faced by other spatial econometric models (Gibbons and Overman, 2012). Furthermore, our approach allows for a more sophisticated incorporation of topographical features, such as roads, rivers, and railways, compared to existing literature (e.g., Allen and Arkolakis, 2014), by constructing additional exogenous regressors informed by the theoretical framework. In the context of quantitative spatial economics, our method requires less information for estimation, representing a significant advantage. Additionally, the proposed methodology demonstrates a strong capability for producing accurate forecasts by accounting for the endogeneity of the spatial dynamics of the variable of interest. This feature makes it particularly useful for forecasting local dynamics of climate change and pollution, as illustrated by Cruz and Rossi-Hansberg (2024).

Our methodology, unlike other spatial econometric models, performs exceptionally well as the frequency and resolution of data increase. Its full potential will be realized with the advent of high-resolution satellite imagery and the involvement of major companies in academic research, which will expand the availability of large, high-resolution georeferenced datasets at higher frequencies (e.g., nightlights, population, climate, and pollution data).[18]

---

[18]For instance, see data from the *Visible Infrared Imaging Radiometer Suite* (VIIRS) on the Suomi satellite launched in 2011 by NASA and the National Oceanic and Atmospheric Administration (NOAA) at https://eogdata.mines.edu/products/vnl/, the Copernicus Programme at https://www.copernicus.eu/en, and *Data For Good*'s High-Resolution Population Density Maps by META at https://dataforgood.facebook.com/dfg/tools.

References

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The economics of density: Evidence from the berlin wall. *Econometrica*, 83(6):2127–2189.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1):223–262.

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Alonso, W. (1964). *Location and land use: toward a general theory of land rent.* Harvard university press.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, 27(3):247–267.

Barthelemy, M. (2016). *The structure and dynamics of cities.* Cambridge University Press.

Benito, J., Urena, F., and Gavete, L. (2001). Influence of several factors in the generalized finite difference method. *Applied Mathematical Modelling*, 25(12):1039–1053.

Blume, L. E., Brock, W. A., Durlauf, S. N., and Jayaraman, R. (2015). Linear social interactions models. *Journal of Political Economy*, 123(2):444–496.

Brock, W. A. and Durlauf, S. N. (2001a). Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235–260.

Brock, W. A. and Durlauf, S. N. (2001b). Interactions-based models. In *Handbook of econometrics*, volume 5, pages 3297–3380. Elsevier.

Brueckner, J. K. (2003). Strategic interaction among governments: An overview of empirical studies. *International regional science review*, 26(2):175–188.

Carrillo, J. A., Craig, K., and Yao, Y. (2019a). Aggregation-diffusion equations: dynamics, asymptotics, and singular limits. In *Active Particles, Volume 2*, pages 65–108. Springer.

Carrillo, J. A., Estrada-Rodriguez, G., Mikolas, L., and Tang, S. (2024). Sparse identification of nonlocal interaction kernels in nonlinear gradient flow equations via partial inversion. *arXiv preprint arXiv:2402.06355*.

Carrillo, J. A., Hittmeir, S., Volzone, B., and Yao, Y. (2019b). Nonlinear aggregation-diffusion equations: radial symmetry and long time asymptotics. *Inventiones mathematicae*, 218(3):889–977.

Catellier, R., D'angelo, Y., and Ricci, C. (2021). A mean-field approach to self-interacting networks, convergence and regularity. *Mathematical Models and Methods in Applied Sciences*, 31(13):2597–2641.

Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics*, volume 5, pages 247–348. Elsevier.

Cressie, N. and Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Cruz, J.-L. and Rossi-Hansberg, E. (2024). The economic geography of global warming. *Review of Economic Studies*, 91(2):899–939.

Desmet, K., Nagy, D. K., and Rossi-Hansberg, E. (2018). The geography of development. *Journal of Political Economy*, 126(3):903–983.

Donaghy, K. P. (2001). Solution and econometric estimation of spatial dynamic models in continuous space and continuous time. *Journal of geographical systems*, 3:257–270.

Duranton, G. and Puga, D. (2020). The economics of urban density. *Journal of Economic Perspectives*, 34(3):3–26.

Duranton, G. and Puga, D. (2023). Urban growth and its aggregate implications. *Econometrica*, 91(6):2219–2259.

Durlauf, S. N., Johnson, P. A., and Temple, J. R. (2005). Growth econometrics. *Handbook of economic growth*, 1:555–677.

Ertur, C. and Koch, W. (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of applied econometrics*, 22(6):1033–1062.

Farlow, S. J. (1993). *Partial differential equations for scientists and engineers*. Courier Corporation.

Fiaschi, D. and Ricci, C. (2023). The spatial evolution of economic activities and the emergence of cities. *arXiv preprint arXiv:2310.07883*.

Fujita, M. and Thisse, J.-F. (1996). Economics of agglomeration. *Journal of the Japanese and international economies*, 10(4):339–378.

Fujita, M. and Thisse, J.-F. (2002). *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge University Press.

Gibbons, S. and Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of regional Science*, 52(2):172–191.

Glaeser, E. L., Kahn, M. E., and Chu, C. (2001). *Job sprawl: Employment location in US metropolitan areas*. Brookings Institution, Center on Urban and Metropolitan Policy Washington, DC.

Graham, C., Kurtz, T. G., Méléard, S., Protter, P. E., Pulvirenti, M., Talay, D., and Méléard, S. (1996). Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. *Probabilistic Models for Nonlinear Partial Differential Equations: Lectures given at the 1st Session of the Centro Internazionale Matematico Estivo (CIME) held in Montecatini Terme, Italy, May 22–30, 1995*, pages 42–95.

Huang, H., Liu, J.-G., and Lu, J. (2019). Learning interacting particle systems: Diffusion parameter estimation for aggregation equations. *Mathematical Models and Methods in Applied Sciences*, 29(01):1–29.

ISTAT (2017). Forme, Livelli e Dinamiche dell'Urbanizzazione in Italia. `www.istat.it/it/files/2017/05/urbanizzazione.pdf`.

Jensen, P. S. (1972). Finite difference techniques for variable grids. *Computers & Structures*, 2(1-2):17–29.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17:99–121.

Kelejian, H. H., Prucha, I. R., and Yuzefovich, Y. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. In *Spatial and spatiotemporal econometrics*, volume 18, pages 163–198. Emerald Group Publishing Limited.

Kneebone, E. (2009). *Job sprawl revisited: The changing geography of metropolitan employment*. Metropolitan Policy Program, Brookings Institution.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, 99(3):483–499.

Krugman, P. (1994). Complex landscapes in economic geography. *The American Economic Review*, 84(2):412–416.

Krugman, P. (1998). Space: the final frontier. *Journal of Economic perspectives*, 12(2):161–174.

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

LeSage, J. P. and Pace, R. K. (2011). Pitfalls in higher order model extensions of basic spatial regression methodology. *Review of Regional Studies*, 41(1):13–26.

Mendes, D. A. and Mendes, V. M. (2015). Parametric models in spatial econometrics: a survey. In *Complexity and Geographical Economics*, pages 51–71. Springer.

Moretti, E. (2004). Human capital externalities in cities. *Handbook of Regional and Urban Economics*, 4:2243–2291.

Oud, J. H., Folmer, H., Patuelli, R., and Nijkamp, P. (2012). Continuous-time modeling with spatial dependence. *Geographical Analysis*, 44(1):29–46.

Piras, G., Donaghy, K. P., and Arbia, G. (2007). Nonlinear regional economic dynamics: continuous-time specification, estimation and stability analysis. *Journal of geographical systems*, 9:311–344.

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9:21–58.

Rosenthal, S. S. and Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2):373–389.

Rossi-Hansberg, E. (2019). Geography of growth and development. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press USA.

Rudin, W. et al. (1964). *Principles of mathematical analysis*, volume 3. McGraw-hill New York.

Sznitman, A.-S. (1991). Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer.

Taylor, M. E. (1975). Reflection of singularities of solutions to systems of differential equations. *Communications on Pure and Applied Mathematics*, 28(4):457–478.

Thomas, J. W. (2013). *Numerical partial differential equations: finite difference methods*, volume 22. Springer Science & Business Media.

Travisi, C. M., Camagni, R., and Nijkamp, P. (2010). Impacts of urban sprawl and commuting: a modelling study for italy. *Journal of Transport Geography*, 18(3):382–392.

Wozniak, A. (2010). Are college graduates more responsive to distant labor market opportunities? *The Journal of Human Resources*, 45(4):944–970.

Wymer, C. R. (2012). Continuous-tme econometrics of structural models. *Studies in Nonlinear Dynamics & Econometrics*, 16(2).

Xu, X. and Lee, L.-f. (2019). Theoretical foundations for spatial econometric research. *Regional Science and Urban Economics*, 76:2–12.

Appendix

A   The matrices for the correction of discretization over time

In this appendix, we describe how to compute the matrices M_S, M_A, M_R and M_D for the discretization in time of Eq. (1). In Section 3.2.2 we saw the case where $\gamma_S, \gamma_A, \gamma_R = 0$. We now deal with the full case. In the general case Eq. (22) becomes:

$$
\begin{aligned}
\partial_{tt}^2 y(t', z^i) &= \partial_t \left[\partial_t y(t, z^i) \right] \big|_{t=t'} = \\
&= \partial_t a\left(t, z^i\right) \big|_{t=t'} + \varphi \partial_t y(t, z^i) \big|_{t=t'} + \\
&+ \gamma_S \Bigg[\partial_{z_1} \left(\partial_t y(t', z) \partial_{z_1} S(z) \right) + \partial_{z_2} \left(\partial_t y(t', z) \partial_{z_2} S(z) \right) \Bigg] \bigg|_{z=z^i} + \\
&+ \gamma_A \Bigg[\partial_{z_1} \left(\partial_t y(t', z) \partial_{z_1} \left(K_{h_A} * y \right)(t', z) + y(t', z) \partial_{z_1} \left(K_{h_A} * \partial_t y \right)(t', z) \right) + \\
&+ \partial_{z_2} \left(\partial_t y(t', z) \partial_{z_2} \left(K_{h_A} * y \right)(t', z) + y(t', z) \partial_{z_2} \left(K_{h_A} * \partial_t y \right)(t', z) \right) \Bigg] \bigg|_{z=z^i} + \\
&+ \gamma_R \Bigg[\partial_{z_1} \left(\partial_t y(t', z) \partial_{z_1} \left(K_{h_R} * y \right)(t', z) + y(t', z) \partial_{z_1} \left(K_{h_R} * \partial_t y \right)(t', z) \right) + \\
&+ \partial_{z_2} \left(\partial_t y(t', z) \partial_{z_2} \left(K_{h_R} * y \right)(t', z) + y(t', z) \partial_{z_2} \left(K_{h_R} * \partial_t y \right)(t', z) \right) \Bigg] \bigg|_{z=z^i} + \\
&+ \gamma_D \Bigg[\partial_{z_1 z_1} \partial_t y(t', z) + \partial_{z_2 z_2} \partial_t y(t', z) \Bigg] \bigg|_{z=z^i}.
\end{aligned}
\tag{29}
$$

Applying the discretization also in space and assuming $\partial_t y(t, z^i)\big|_{t=t'} \approx \Delta_\tau y_i$ as in Section 3.2.2, we have

$$
\begin{aligned}
\partial_{tt}^2 y(t', z^i) \approx{} & \partial_t a\left(t', z^i\right) + \rho_\varphi \Delta_\tau y_i + \\
& + \rho_S \left[M_{z_1}\left(\Delta_\tau \mathbf{y} \odot M_{z_1}\mathbf{s}\right) + M_{z_2}\left(\Delta_\tau \mathbf{y} \odot M_{z_2}\mathbf{s}\right) \right]_i + \\
& + \rho_A \Big[M_{z_1}\left(\Delta_\tau \mathbf{y} \odot M_{z_1} W_{h_A}\mathbf{y} + \mathbf{y} \odot M_{z_1} W_{h_A}\Delta_\tau \mathbf{y}\right) + \\
& \quad + M_{z_2}\left(\Delta_\tau \mathbf{y} \odot M_{z_2} W_{h_A}\mathbf{y} + \mathbf{y} \odot M_{z_2} W_{h_A}\Delta_\tau \mathbf{y}\right) \Big]_i + \\
& + \rho_R \Big[M_{z_1}\left(\Delta_\tau \mathbf{y} \odot M_{z_1} W_{h_R}\mathbf{y} + \mathbf{y} \odot M_{z_1} W_{h_R}\Delta_\tau \mathbf{y}\right) + \\
& \quad + M_{z_2}\left(\Delta_\tau \mathbf{y} \odot M_{z_2} W_{h_R}\mathbf{y} + \mathbf{y} \odot M_{z_2} W_{h_R}\Delta_\tau \mathbf{y}\right) \Big]_i + \\
& + \rho_D \left[(M_{z_1 z_1} + M_{z_2 z_2})\Delta_\tau \mathbf{y}\right]_i, \quad (30)
\end{aligned}
$$

for some $\rho_\varphi \approx \varphi$, $\rho_S \approx \gamma_S$, $\rho_A \approx \gamma_A$, $\rho_R \approx \gamma_R$ and $\rho_D \approx \gamma_D$.

The expression related to the coefficient ρ_D can be immediately interpreted as

$$
\rho_D \left[(M_{z_1 z_1} + M_{z_2 z_2})\Delta_\tau \mathbf{y}\right]_i = \rho_D (M_D \Delta_\tau \mathbf{y})_i
$$

by introducing $M_D \equiv M_{z_1 z_1} + M_{z_2 z_2}$. The terms related to the coefficients ρ_S, ρ_A and ρ_R, even though cannot be immediately expressed in matrix form, are linear with respect to $\Delta_\tau \mathbf{y}$. Therefore it is possible to construct matrices M_S, M_A and M_R depending on $\mathbf{y}$ such that Eq. (30) becomes

$$
\begin{aligned}
\partial_{tt}^2 y(t', z^i) \approx{} & \partial_t a\left(t', z^i\right) + \rho_\varphi \Delta_\tau y_i + \\
& + \rho_S (M_S \Delta_\tau \mathbf{y})_i + \rho_A (M_A \Delta_\tau \mathbf{y})_i + \rho_R (M_R \Delta_\tau \mathbf{y})_i + \rho_D (M_D \Delta_\tau \mathbf{y})_i.
\end{aligned}
$$

To construct such matrices recall that given any linear function $\psi : \mathbb{R}^N \to \mathbb{R}^N$, this can be represented in matrix form with respect to the canonical basis of $\mathbb{R}^N$ $\{\mathbf{e}_i\}_{i=1,\dots,N}$ by the matrix which has as the i-th row the vector $\psi(\mathbf{e}_i)$.

Eq. (31) is the analogous of Eq. (23) when all the terms of Eq. (1) are present. Repeating the same computation of Section 3.2.2 we end up with the analogous of

Eq. (24) with all the terms, which reads

$$
\begin{aligned}
\Delta_\tau \mathbf{y} \approx & \left[\left(1 - \frac{\tau \rho_\varphi}{2}\right) \mathbf{I} - \frac{\tau \rho_S}{2} M_S - \frac{\tau \rho_A}{2} M_A - \frac{\tau \rho_R}{2} M_R - \frac{\tau \rho_D}{2} M_D \right]^{-1} \times \\
\times & \left[\mathbf{a} + \frac{\tau}{2} \Delta_\tau \mathbf{a} + \varphi \mathbf{y} + \right. \\
+ & \gamma_S \left[M_{z_1} \left(\mathbf{y} \odot M_{z_1} \mathbf{s} \right) + M_{z_2} \left(\mathbf{y} \odot M_{z_2} \mathbf{s} \right) \right] + \\
+ & \gamma_A \left[M_{z_1} \left(\mathbf{y} \odot M_{z_1} W_{h_A} \mathbf{y} \right) + M_{z_2} \left(\mathbf{y} \odot M_{z_2} W_{h_A} \mathbf{y} \right) \right] + \\
+ & \gamma_R \left[M_{z_1} \left(\mathbf{y} \odot M_{z_1} W_{h_R} \mathbf{y} \right) + M_{z_2} \left(\mathbf{y} \odot M_{z_2} W_{h_R} \mathbf{y} \right) \right] + \\
+ & \left. \gamma_D (M_{z_1 z_1} + M_{z_2 z_2}) \mathbf{y} \right].
\end{aligned}
\tag{31}
$$

From Eq. (31) by rearranging the terms we arrive at Eq. (26).

B   Estimation procedure of spatial matrix W_ϵ

In this section, we describe the data-driven procedure for the calculation of W_ϵ. Let $\hat{\mathbf{e}}$ be the $1 \times N$ vector of residuals of Model (4) estimated by ML without controlling for spatial correlation in the errors, and W_q the spatial matrix based on the q-th order of contiguity across cells and zero diagonal. As the first step, estimate by OLS the model:

$$
\hat{\mathbf{e}} = \left[\ell_1 W_1 + \ell_2 \left(W_2 - W_1 \right) + \ell_3 \left(W_3 - W_2 \right) + \cdots + \ell_Q \left(W_Q - W_{Q-1} \right) \right] \hat{\mathbf{e}} + \mathbf{u}, \tag{32}
$$

where $\mathbf{u}$ is a vector of random components. Then, calculate the spatial weight matrix W_ϵ used in Model (4) with the specification of error in Eq. (10) as:

$$
W_\epsilon = \hat{\ell}_1 W_1 + \hat{\ell}_2 \left(W_2 - W_1 \right) + \hat{\ell}_3 \left(W_3 - W_2 \right) + \cdots + \hat{\ell}_{\hat{Q}} \left(W_{\hat{Q}} - W_{\hat{Q}-1} \right), \tag{33}
$$

where $\hat{Q}$ is the appropriate maximum order of contiguity suggested by the statistical significance of $\hat{\ell}$'s. As regards the desired properties, W_ϵ has zero diagonal, is symmetric, and allows for nonlinear diffusion of errors over space; in particular, the intensity of diffusion of the first-contiguous cells is equal to $\hat{\ell}_1$, of the only second-contiguous cells is equal to $\hat{\ell}_2$, etc... The use of W_ϵ of Eq. 33 implies the estimated λ should be very close to one. However, this could not be the case because, in principle, OLS estimates of Model (32) are biased for the presence of endogeneity. More sophisticated methods of estimate, such as ML, could overcome such bias but also imply not trivial numerical difficulties which we leave to future research (LeSage and Pace, 2011).

Online Appendix

C   The micro-foundation of the spatial growth model

This appendix summarizes the main ideas about the microfoundation of Eq. (1). Consider a set of N_a rational agents, which can be understood as units of production/consumers. For the sake of simplicity, we briefly describe here the case where $\varphi = 0$ and $a(t,z) = 0 \,\forall (t,z)$, i.e. there is no change in the total number of agents but they only relocate across space. The case $\varphi \neq 0$ can be treated similarly, see for example Catellier et al. (2021). Each agent is identified by index i, with $i = 1, \ldots, N_a$, and is characterised by its location in the domain $\Omega \subseteq \mathbb{R}^2$, labelled by X_t^{i,N_a}. At $t = 0$ agents are independently distributed at random in the domain Ω following a common probability density distribution on Ω denoted by $y_0(z)$. For $t > 0$ each agent's location evolves by obeying the following Stochastic Differential Equation (SDE):

$$
\begin{aligned}
dX_t^{i,N_a} = & - \gamma_S \nabla_z S\left(X_t^{i,N_a}\right) dt + \\
& - \gamma_A \frac{1}{N_a} \sum_{j=1}^{N_a} \nabla_z K_{h_A}\left(X_t^{i,N_a} - X_t^{j,N_a}\right) dt + \\
& - \gamma_R \frac{1}{N_a} \sum_{j=1}^{N_a} \nabla_z K_{h_R}\left(X_t^{i,N_a} - X_t^{j,N_a}\right) dt + \\
& + \sqrt{2\gamma_D}\, dB_t^i,
\end{aligned}
\tag{34}
$$

where $(B_t^i)_{i \in \mathbb{N}}$ is a sequence of independent Brownian motions, and $\gamma_A < 0$, while γ_S, γ_R and $\gamma_D > 0$, to respect the coherence with the phenomena (spatial non-uniformity, aggregation, repulsion, diffusion) we are interested to model.

According to Eq. (34) the spatial position of each agent is evolving by keeping into account its relative position with respect to all other agents. In particular, the first term on the right-hand side of Eq. (34) expresses the tendency of agents to move where the function S is lower, which can be decided by the agents' spatial distribution and/or the particular characteristics of different locations. The second and third terms reflect the interactions among agents. In particular, by the linearity

of the derivative:

$$\frac{1}{N_a} \sum_{j=1}^{N_a} \nabla_z K_{h_A} \left(X_t^{i,N_a} - X_t^{j,N_a} \right) = \nabla_z \left(\frac{1}{N_a} \sum_{j=1}^{N_a} K_{h_A} \left(\cdot - X_t^{j,N_a} \right) \right) (X_t^{i,N_a}); \text{ and}$$

$$\frac{1}{N_a} \sum_{j=1}^{N_a} \nabla_z K_{h_R} \left(X_t^{i,N_a} - X_t^{j,N_a} \right) = \nabla_z \left(\frac{1}{N_a} \sum_{j=1}^{N_a} K_{h_R} \left(\cdot - X_t^{j,N_a} \right) \right) (X_t^{i,N_a}),$$

that is, each agent is moving along the direction of the gradient of a local average (the concept of locality is defined by the functions K_{h_A} and K_{h_R}) of the other agents' location. Finally, the fourth term of Eq. (34) represents the agents' *idiosyncratic and independent random movements* and is expressed through the independent additive noise B_t^i.

The *empirical distribution* of agent's location is:

$$E_t^{N_a} := \frac{1}{N_a} \sum_{i=1}^{N_a} \delta_{X_t^{i,N_a}}, \tag{35}$$

where δ_z is the random variable on $\mathbb{R}^2$ with unitary mass in the point z. $E_t^{N_a}$ is a continuous set of random variables on $\mathbb{R}^2$ depending on time. For any given $N_a \in \mathbb{N}$ and $t > 0$, this distribution is singular, in the sense that is a distribution over $\mathbb{R}^2$ which does not admit a probability density function, since it has a positive probability only on a finite set (corresponding to the location of the N_a agents). However, when N_a goes to infinity the family of random variable $E_t^{N_a}$ becomes diffuse and converges (in distribution) to a continuous family of random variables over $\mathbb{R}^2$, labelled by E_t for any $t > 0$. The distribution of E_t is now regular and admits a probability density function for every t, called $y(t, z)$. An explicit expression for $y(t, z)$ for every t is not available. However one can prove that the probability density function $y(t, z)$ is the unique solution to Eq. (1) (see Sznitman, 1991, Section 1.1).

D   A numerical investigation of the properties of the spatial growth model

An explicit solution for Eq. (1) is not available in general. We will then proceed to discuss some of the basic properties of Eq. (1) by numerical simulations. In all examples, we make some simplifications to better discern the peculiar features of the model from the exogenous effects related to the particular case considered in Section 4. In particular, we take the total cross-sectional amount of the variable y of the system constant and equal to 1, i.e. we take $\varphi = 0$. Moreover, we take the exogenous function $S(z)$ to be identically zero. We also neglect the repulsive effect, i.e. set $\gamma_R = 0$. Summarising, we consider only the *aggregation* (γ_A) and *diffusion* (γ_D) effects. We study the problem on a squared domain $\Omega = [0,4] \times [0,4]$ with a discrete set of locations uniformly spaced with a distance between contiguous locations of $\Delta z = 10^{-2}$ (i.e. 160,000 total points).

In figures 11 and 12 we set an initial condition that is flat at the centre of the domain and zero close to the boundary, with some intermediate values in between so that the initial profile is a continuous function. We then set $\gamma_A = -0.01, \gamma_D = 0.005$ and explore the impacts of changes in h_A since the distance at which aggregation takes place is one of the main characteristics of the system which determines the evolution of the spatial pattern. Figure 11 reports the distribution dynamics of the baseline model which highlights the dynamics of aggregation of income, i.e. the emergence of a city, in a central location at around $t = 20$. We notice that there is an initial temporary formation of four smaller clusters around $t = 5$, which then agglomerates together to contribute to the formation of the single large cluster in the centre of the domain. In Figure 12 we keep the initial condition unchanged and only reduce the distance of aggregation from 0.4 to 0.3. The spatial pattern in the first few periods is roughly the same, exhibiting the formation of the four temporary clusters around period $t = 5$. However, since the distance at which aggregation takes place is smaller, the four clusters are now so far apart that they are not able to merge anymore. Therefore we observe at the final period $t = 20$ a configuration which is made of four separate clusters instead of a single larger one. The spatial pattern strongly resembles the ones discussed in Krugman (1994) and in Barthelemy (2016, cap. 8).

47

E The computation of the matrices of spatial partial derivative

This appendix introduces to *generalized finite difference method* for calculating the spatial partial derivatives when observations are not equally distributed across the space, as originally introduced in Jensen (1972) (for a detailed overview of the method see, e.g., Benito et al., 2001).

Let $y(z)$ be the value of a function at location $z \in \mathbb{R}^2$; define a set of neighbouring areas z^j whose function values are indicated by $y(z^j)$, for $j = 1, \cdots, n_s$, where n_s is the number of neighbouring locations. Define the following function:

$$
\begin{aligned}
B(z) \ = \ & \sum_{j=1}^{n_s} \left\{ \left[y(z) - y(z^j) + h_j \frac{\partial y(z)}{\partial z_1} + k_j \frac{\partial y(z)}{\partial z_2} + \right.\right. \\
& \left.\left. + \ \frac{1}{2} \left(h_j^2 \frac{\partial^2 y(z)}{\partial z_1^2} + k_j^2 \frac{\partial^2 y(z)}{\partial z_2^2} + 2 h_j k_j \frac{\partial^2 y(z)}{\partial z_1 \partial z_2} \right) \right] w(h_j, k_j) \right\}^2 , \quad (36)
\end{aligned}
$$

where $h_j = z_1 - z_1^j$, $k_j = z_2 - z_2^j$ and $w(\cdot, \cdot)$ is a weighting function decreasing in both arguments and always not negative. The function $B(z)$ is a weighted linear combination of squares of the error that one has by approximating the function $y(z^j)$ with its second-order linear approximation in the point $y(z)$, for every $j = 1, \ldots, n_s$. Therefore, if this approximation is sufficiently accurate, $B(z)$ is close to zero. Given $y(z)$, $y(z^j)$, h_j, k_j and w_j, for $j = 1, 2, \cdots, n_s$, Eq. (36) allows to calculate the values of $\frac{\partial y(z)}{\partial z_1}$, $\frac{\partial y(z)}{\partial z_2}$, $\frac{\partial^2 y(z)}{\partial z_1^2}$, $\frac{\partial^2 y(z)}{\partial z_2^2}$ and $\frac{\partial^2 y(z)}{\partial z_1 \partial z_2}$ under the condition that $B(z)$ is minimized. In particular, due to the quadratic shape of the function B, this minimization amounts to solving the following system of linear equation (see Benito et al., 2001, p. 6):

$$
D_z \ = \ \begin{bmatrix} \dfrac{\partial y(z)}{\partial z_1} \\ \dfrac{\partial y(z)}{\partial z_2} \\ \dfrac{\partial^2 y(z)}{\partial z_1^2} \\ \dfrac{\partial^2 y(z)}{\partial z_2^2} \\ \dfrac{\partial^2 y(z)}{\partial z_1 \partial z_2} \end{bmatrix} = \mathbf{D} \begin{bmatrix} y(z) \\ y(z^1) \\ y(z^2) \\ \cdots \\ y(z^{n_s}) \end{bmatrix} , \quad (37)
$$

with $\mathbf{D} \equiv \mathbf{A}^{-1} \mathbf{B}$ is a $(5 \times (n_s + 1))$ matrix, $w_j = w(h_j, k_j)$ and where $\mathbf{A}$ is a

symmetric (5×5) matrix defined as:

$$
A = \begin{bmatrix}
\sum_{j=1}^{n_s} h_j^2 w_j^2 & \sum_{j=1}^{n_s} h_j k_j w_j^2 & \frac{1}{2}\sum_{j=1}^{n_s} h_j^3 w_j^2 & \frac{1}{2}\sum_{j=1}^{n_s} h_j k_j^2 w_j^2 & \sum_{j=1}^{n_s} h_j^2 k_j w_j^2 \\
 & \sum_{j=1}^{n_s} k_j^2 w_j^2 & \frac{1}{2}\sum_{j=1}^{n_s} h_j^2 k_j w_j^2 & \frac{1}{2}\sum_{j=1}^{n_s} k_j^3 w_j^2 & \sum_{j=1}^{n_s} h_j k_j^2 w_j^2 \\
 & & \frac{1}{4}\sum_{j=1}^{n_s} h_j^4 w_j^2 & \frac{1}{4}\sum_{j=1}^{n_s} h_j^2 k_j^2 w_j^2 & \frac{1}{2}\sum_{j=1}^{n_s} h_j^3 k_j w_j^2 \\
 & & & \frac{1}{4}\sum_{j=1}^{n_s} k_j^4 w_j^2 & \frac{1}{2}\sum_{j=1}^{n_s} h_j k_j^3 w_j^2 \\
 & & & & \sum_{j=1}^{n_s} h_j^2 k_j^2 w_j^2
\end{bmatrix} \tag{38}
$$

and $\mathbf{B}$ is a $(5 \times (n_s + 1))$ matrix defined as:

$$
B = \begin{bmatrix}
-\sum_{j=1}^{n_s} h_j w_j^2 & h_1 w_1^2 & h_2 w_2^2 & \cdots & h_{n_s} w_{n_s}^2 \\
-\sum_{j=1}^{n_s} k_j w_j^2 & k_1 w_1^2 & k_2 w_2^2 & \cdots & k_{n_s} w_{n_s}^2 \\
-\frac{1}{2}\sum_{j=1}^{n_s} h_j^2 w_j^2 & \frac{1}{2}h_1^2 2 w_1^2 & \frac{1}{2}k_1^2 w_2^2 & \cdots & \frac{1}{2}h_{n_s}^2 w_{n_s}^2 \\
-\frac{1}{2}\sum_{j=1}^{n_s} k_j^2 w_j^2 & \frac{1}{2}h_2^2 w_1^2 & \frac{1}{2}k_2^2 w_2^2 & \cdots & \frac{1}{2}k_{n_s}^2 w_{n_s}^2 \\
-\sum_{j=1}^{n_s} h_j k_j w_j^2 & h_1 k_1 w_1^2 & h_2 k_2 w_2^2 & \cdots & h_{n_s} k_{n_s} w_{n_s}^2
\end{bmatrix} . \tag{39}
$$

Benito et al. (2001) suggests using as a weighting function:

$$
w(d_j) = 1 - 6\left(\frac{d_j}{dm_j}\right)^2 + 8\left(\frac{d_j}{dm_j}\right)^3 - 3\left(\frac{d_j}{dm_j}\right)^4 , \tag{40}
$$

where $d_j^2 = h_j^2 + k_j^2$ is the squared distance between z and z_j and dm_j is the maximum distandiffusivece on all possible neighbouring locations.

The set of all neighbouring locations used to approximate the partial derivatives in location z is called the *star* of the location; the choice of the elements of the *star* is an important factor to ensure the accuracy of the method. We will adopt the *closest neighbourhood criterion*, that is, fixed the number of elements of the *star* n_s, we select the n_s closest locations to each given location.

For every given location z, it is, therefore, possible to build the matrix D and all the partial derivatives in location z up to the second order. To carry out these operations set $(z^i)_{i=1,\dots,N}$ a finite number of locations; for each of these locations denote by $\mathbf{J^i} = (\mathbf{J^i}_k)_{k=1,\dots,n_s}$ the vector of indices of locations in the *star* of location z^i (by definition of star, i is not an element of $\mathbf{J^i}$), and by $\mathbf{D_i}$ the matrix

$\mathbf{D}$ associated to the location z^i; thus

$$
\begin{bmatrix}
\dfrac{\partial y(z^i)}{\partial z_1} \\[4pt]
\dfrac{\partial y(z^i)}{\partial z_2} \\[4pt]
\dfrac{\partial^2 y(z^i)}{\partial z_1^2} \\[4pt]
\dfrac{\partial^2 y(z^i)}{\partial z_2^2} \\[4pt]
\dfrac{\partial^2 y(z^i)}{\partial z_1 \partial z_2}
\end{bmatrix}
= \mathbf{D_i}
\underbrace{
\begin{bmatrix}
y(z^i) \\
y(z^{\mathbf{J^i}_1}) \\
y(z^{\mathbf{J^i}_2}) \\
\vdots \\
y(z^{\mathbf{J^i}_{n_s}})
\end{bmatrix}
}_{\overline{\mathbf{y}}_i}
= \mathbf{D_i}\overline{\mathbf{y}}_i.
\tag{41}
$$

Therefore, denoting by $(D_i)_{k,}$ the k-th row of the matrix D_i

$$
\begin{aligned}
\frac{\partial y(z^i)}{\partial z_1} &= \langle (\mathbf{D_i})_{1,}, \overline{\mathbf{y}}_i \rangle \, ; \\[6pt]
\frac{\partial y(z^i)}{\partial z_2} &= \langle (\mathbf{D_i})_{2,}, \overline{\mathbf{y}}_i \rangle \, ; \\[6pt]
\frac{\partial^2 y(z^i)}{\partial z_1^2} &= \langle (\mathbf{D_i})_{3,}, \overline{\mathbf{y}}_i \rangle \, ; \text{ and} \\[6pt]
\frac{\partial^2 y(z^i)}{\partial z_2^2} &= \langle (\mathbf{D_i})_{4,}, \overline{\mathbf{y}}_i \rangle \, ,
\end{aligned}
$$

where $\langle \cdot, \cdot \rangle$ is the standard scalar product of $\mathbb{R}^{n_s+1}$.

To compute all the partial derivatives in all the locations via single matrix multiplications, set

$$
\mathbf{y} = (y(z^1), y(z^2), \ldots, y(z^N))^t
$$

and the vectors of partial derivatives in all the locations as

$$
\frac{\partial y}{\partial z_1} =
\begin{bmatrix}
\dfrac{\partial y(z^1)}{\partial z_1} \\[4pt]
\dfrac{\partial y(z^2)}{\partial z_1} \\[4pt]
\vdots \\[4pt]
\dfrac{\partial y(z^N)}{\partial z_1}
\end{bmatrix},
\frac{\partial y}{\partial z_2} =
\begin{bmatrix}
\dfrac{\partial y(z^1)}{\partial z_2} \\[4pt]
\dfrac{\partial y(z^2)}{\partial z_2} \\[4pt]
\vdots \\[4pt]
\dfrac{\partial y(z^N)}{\partial z_2}
\end{bmatrix},
\frac{\partial^2 y}{\partial z_1^2} =
\begin{bmatrix}
\dfrac{\partial^2 y(z^1)}{\partial z_1^2} \\[4pt]
\dfrac{\partial^2 y(z^2)}{\partial z_1^2} \\[4pt]
\vdots \\[4pt]
\dfrac{\partial^2 y(z^N)}{\partial z_1^2}
\end{bmatrix},
\frac{\partial^2 y}{\partial z_2^2} =
\begin{bmatrix}
\dfrac{\partial^2 y(z^1)}{\partial z_2^2} \\[4pt]
\dfrac{\partial^2 y(z^2)}{\partial z_2^2} \\[4pt]
\vdots \\[4pt]
\dfrac{\partial^2 y(z^N)}{\partial z_2^2}
\end{bmatrix}.
$$

Consider the $(N \times N)$ matrices M_{z_1}, M_{z_2}, $M_{z_1 z_1}$ and $M_{z_2 z_2}$ defined by

$$
(M_{z_1})_{i,j} = \begin{cases} (\mathbf{D_i})_{1,1} & \text{if } j = i, \\ (\mathbf{D_i})_{1,k+1} & \text{if } j = \mathbf{J^i}_k \text{ for some } k \in \{1, \ldots, n_s\}, \\ 0 & \text{otherwise,} \end{cases}
$$

$$
(M_{z_2})_{i,j} = \begin{cases} (\mathbf{D_i})_{2,2} & \text{if } j = i, \\ (\mathbf{D_i})_{2,k+1} & \text{if } j = \mathbf{J^i}_k \text{ for some } k \in \{1, \ldots, n_s\}, \\ 0 & \text{otherwise,} \end{cases}
$$

$$
(M_{z_1 z_1})_{i,j} = \begin{cases} (\mathbf{D_i})_{3,3} & \text{if } j = i, \\ (\mathbf{D_i})_{3,k+1} & \text{if } j = \mathbf{J^i}_k \text{ for some } k \in \{1, \ldots, n_s\}, \\ 0 & \text{otherwise,} \end{cases}
$$

$$
(M_{z_2 z_2})_{i,j} = \begin{cases} (\mathbf{D_i})_{4,4} & \text{if } j = i, \\ (\mathbf{D_i})_{4,k+1} & \text{if } j = \mathbf{J^i}_k \text{ for some } k \in \{1, \ldots, n_s\}, \\ 0 & \text{otherwise,} \end{cases}
$$

then, for all $i = 1, \ldots, N$,

$$
\frac{\partial y(z^i)}{\partial z_1} \approx (M_{z_1} \mathbf{y})_i, \quad \frac{\partial y(z^i)}{\partial z_2} \approx (M_{z_2} \mathbf{y})_i, \quad \frac{\partial^2 y(z^i)}{\partial z_1^2} \approx (M_{z_1 z_1} \mathbf{y})_i, \quad \frac{\partial^2 y(z^i)}{\partial z_2^2} \approx (M_{z_2 z_2} \mathbf{y})_i.
$$

G   The estimated W_ϵ for Italian municipalities

Table 5 reports the estimated $\hat{\ell}$'s of Model (32) in Appendix B for Italian municipalities. The most appropriate maximum order of contiguity is $\hat{Q} = 5$, decided based on the statistical significance of the estimated parameters. Notice that spatial correlation appears to suddenly drop after the first order of contiguity.

	Estimate	Std. Error	t value	Pr(>|t|)
ℓ_1	0.1284	0.0035	36.56	0.0000
ℓ_2	0.0043	0.0022	2.01	0.0448
ℓ_3	0.0070	0.0017	4.01	0.0001
ℓ_4	0.0076	0.0016	4.88	0.0000
ℓ_5	0.0045	0.0014	3.14	0.0017
ℓ_6	-0.0004	0.0013	-0.27	0.7872
ℓ_7	0.0015	0.0012	1.26	0.2059
ℓ_8	0.0016	0.0011	1.37	0.1696
ℓ_9	0.0006	0.0011	0.52	0.6003
ℓ_{10}	0.0015	0.0011	1.36	0.1748

Table 5: The estimate of Model (32)'s coefficients in Appendix B with $Q = 10$. The calculation of W_ϵ is based on the first five coefficients, i.e. $Q = 5$, decided based on their statistical significance.