Shrinkage in Space —
Spillovers in a Bayesian Hierarchical Model

Nikolas Kuschnig*

Preliminary Draft

February 27, 2023

Abstract

In this paper, I present a modelling approach to jointly investigate connectivity between observations and its consequences — spillover effects. The approach is fully Bayesian, and uses hierarchical shrinkage priors to flexibly provide regularization where needed and let the data speak where it is possible. I make the prior information that is embodied in the restrictive assumptions of previous spatial models explicit, and loosen them by estimating connectivity parameters. For effective estimation, I develop efficient sampling procedures and a Gaussian process approximation to evaluate Jacobian determinants.

Keywords: shrinkage prior, spatial, network, neighborhood, peer effects

1 Introduction

Economic theory frequently suggests connections between individual units of observation. Connectivity and the resulting spillover effects between units lie at the heart of pressing research questions in microeconomics (Ambrus et al., 2014; Chyn and Katz, 2021; Ioannides and Datcher Loury, 2004; Weidmann and Deming, 2021) and macroeconomics (Acemoglu et al., 2016; Crespo Cuaresma et al., 2019; Gofman, 2017; Rose, 2004). Good answers to these questions and deeper insights can be important for education policy (Board and Meyer-ter Vehn, 2021; Lin, 2010; Mele, 2020), labor market outcomes (Beaman, 2012; Hensvik and Skans, 2016; Munshi, 2003), supply chain management (Acemoglu et al., 2012; Atalay et al., 2014, 2011; Kranton and Minehart, 2001), innovation (Bloom et al., 2013; Ductor et al., 2014; König et al., 2019; Newman, 2001). However, empirical studies of spillover effects abstract from the nature of connections and modelling approaches are limited. Current econometric methods for analyzing spillover effects generally rely on holding the structure of connectivity fixed, thus obscuring uncertainties and potentially distorting results.

This paper introduces an integrated model to jointly analyze both the structure and consequences of connectivity between units. I propose a Bayesian hierarchical approach to comprehensively address both issues. Suitable shrinkage priors allow for more nuanced assumptions; regularization is flexibly imposed where needed, while important aspects of the model are freed up and learned from the data. The resulting model facilitates the explicit treatment of connectivity structures and spillovers in a layered framework. As I demonstrate with an application to deforestation spillovers from croplands in the Brazilian Amazon, this allows us to formulate models that better reflect reality, and capture the uncertainties surrounding it.

The main contribution of this paper is a framework for jointly estimating spillover (or peer) effects and connectivity structures, i.e. spatial weight or adjacency matrices. A number of earlier studies pursue similar approaches in the spatial econometric (Debarsy and LeSage, 2020; Lam and Souza, 2020; Qu and Lee, 2015; Zhang and Yu, 2018) and the network econometric (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Johnsson and Moon, 2021) literature. These studies generally adopt model averaging approaches, and are limited to few discrete candidates for the connectivity structure. The hierarchical setup that I propose in this paper eliminates this constraint. It provides us with a flexible and nuanced way of modelling and estimating more general forms of connectivity, conveying a more comprehensive picture of spillover effects. For this approach to work in practice, there

are three main obstacles to overcome.

The first challenge lies in the forms of the connectivity structure themselves. Here, this paper focuses on common forms of connectivity used in the spatial econometric literature. I decompose the spatial weight matrix to highlight simplifying assumptions made, identify implicit parameters for estimation, and gather insights for suitable prior distributions. For the underlying connectivity structure, I investigate two distance-based specifications in detail. These are nearest-neighbor and distance-decay specifications, with functional forms revolving around a discrete and a continuous parameter. Lastly, I derive conclusive parameter bounds for the overall connectivity strength (in the autoregressive terms) that ensure non-singularity and stationarity of the model.

The second challenge are the prior distributions themselves. In the current literature, priors are arguably neglected. Uniform priors are used most commonly, in an attempt to feign ignorance regarding specific values, but distort other important dimensions of the prior, such as the size. Other priors are limited in their flexibility, and cannot accommodate the nuanced prior information available. In the paper, I elaborate on these nuances, discuss potential sources of prior information, and introduce prior distributions that can reflect this information. I propose a Beta-Gamma mixture prior for the (autoregressive) connectivity strength that can accommodate flexible shapes, providing sensible regularization without distorting estimates. This facilitates the estimation of structural connectivity parameters that have previously been fixed — I discuss suitable priors for modelling the most prominent parameters.

The third challenge is of a technical nature and concerns estimation of the model, and the computations involved. Posterior inference relies on Markov chain Monte Carlo (MCMC) sampling or variational methods, and the interdependence that results from connectivity can make even simple models computationally prohibitive. In particular, models with an autoregressive term rely on the evaluation of a costly Jacobian determinant. I focus on full posterior inference, and develop efficient sampling schemes for the hierarchical and structural parameters, which facilitate straightforward estimation of extensible models. For such models, the Jacobian determinant features at least one additional dimension, making established procedures obsolete. In order to still allow for rapid and accurate estimation, I develop an adaptive Gaussian process approximation. These optimizations allow for full Bayesian inference in spatial econometric models that can be extended with little overhead.

The remainder of this paper is structured as follows.

2 Framework

In this section, I (i) provide a brief introduction into spatial econometric models, (ii) discuss and decompose the connectivity between observations, (iii) derive conclusive bounds for the overall connectivity strength, and (iv) discuss the interpretation of spatial econometric models.

2.1 Methods

The comprehensive spatial econometric model is given by

$$\mathbf{y} = \alpha + \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{e}, \qquad \mathbf{e} = \rho \mathbf{W}\mathbf{e} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ is a vector of responses, $\mathbf{X} \in \mathbb{R}^{n \times p}$ is a matrix of explanatories, and $\boldsymbol{\varepsilon} \sim \mathrm{N}(0, \sigma^2 \mathbf{I})$ is a vector of Gaussian errors with spherical variance. The connectivity structure tells us which units are connected, is assumed known, and is encoded in the connectivity matrix $\mathbf{W} \in \mathbb{R}^{n \times n}$, with element $w_{ij} > 0$ if i and j ($i \neq j$) are neighbors, and 0 otherwise.[1] Pre-multiplication of a vector with $\mathbf{W}$ generates a weighted average of neighboring values, also referred to as spatial lag. The standard parameters of this model are α, the intercept, $\boldsymbol{\beta} \in \mathbb{R}^p$, the regression coefficients, and σ^2, the variance of the error term. The spatial parameters are λ, the spatially autoregressive parameter, $\boldsymbol{\theta} \in \mathbb{R}^p$, the regression coefficients of the spatially lagged explanatories, and ρ, the spatial autocorrelation parameter.

We are interested in modelling the connectivity between units. For simplicity, and to cover interesting constraints in terms of singularity, stationarity, and computation, we will focus our attention on the spatially autoregressive term. Notice, that we can express this model using a latent variable as

$$\mathbf{z} = \alpha + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

$$\mathbf{y} = \mathbf{S}(\lambda \mid \mathbf{W})^{-1} \mathbf{z}, \tag{3}$$

where $\mathbf{z}$ is the latent variable, and $\mathbf{S} = (\mathbf{I} - \lambda \mathbf{W})$ can be interpreted as a connectivity filter. With this formulation, the nested linear model in Equation 2 is apparent, allowing us to concentrate on the filter and the connectivity between units.

---

[1]The connectivity matrix corresponds to a (weighted) graph (or network) with no self-loops.

2.1.1 Connectivity structure

For our hierarchical modelling approach, it helps to decompose the connectivity into its parts. Based on applied practice in the spatial econometric literature, we re-express the connectivity in the following form:

$$f(\lambda \mid \mathbf{W}) = \lambda \mathbf{W} = \lambda \operatorname{diag}(\boldsymbol{\xi}) \, \boldsymbol{\Psi}, \tag{4}$$

where λ determines the overall strength of connectivity, $\boldsymbol{\xi} \in \mathbb{R}^n$ determines the strength of incoming connectivity for individual units, and $\boldsymbol{\Psi} \in \mathbb{R}^{n \times n}$ is an adjacency matrix that holds the connectivity structure, indicating which units are neighbors. Of these three elements, only λ is estimated in the literature; the other elements are assumed to be known and fixed at a convenient value that is (implicitly) supported by prior information.

The adjacency matrix, $\boldsymbol{\Psi}$, stems from a function g that is often based on the distances between units (i.e. a metric d on the locations $\mathcal{L}$). Notable exceptions include functions based on the incidence or length of shared borders. The most common types of distances are Euclidean or spherical distances between geographic locations, although ones based on travel times, transport costs, or trade volumes also see some use. Distance matrices themselves have some theoretically and practically undesirable properties, such as a lack of sparsity,[2] and adjacency matrices are generally constructed using a derived function. One instance is the *kŽ VŠoNef ` VⱮZTad Xg` Uf[a`*, defined as

$$[\boldsymbol{\Psi}]_{ij} = g^{kn}(k \mid d_{ij}) = \begin{cases} 1, & \text{if } d_{ij} \leq \kappa(k \mid d_i), \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

where $k \in \mathbb{N}$ is the number of neighbors, the scalar $d_{ij} = d(i, j)$ is a shorthand for the distance between units i and j, and $\kappa(k \mid d_i)$ yields the kth smallest distance from unit i.

Another example is the *[` hⱮteVⱮ[efS` UⱮⱮUSk Xg` Uf[a`*, defined as

$$[\boldsymbol{\Psi}]_{ij} = g^{dd}(\delta \mid d_{ij}) = \begin{cases} d_{ij}^{-\delta}, & \text{if } i \neq j \text{ and } d_{ij} \leq d^{\max}, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

where $\delta \in \mathbb{R}^+$ determines the speed of decay. Additional parameters, such as $d^{\max}$, which is a distance threshold that induces sparsity, are possible. In contrast to Equation 5, this function yields symmetric adjacency matrices, and, due to the continuous parameter, it

---

[2]For any distance we have (i) $d(a, b) = 0 \Leftrightarrow a = b$, (ii) $d(a, b) = d(b, a)$, and (iii) $d(a, c) \leq d(a, b) + d(b, c)$, and, as a result, no sparsity — i.e. $d(a, b) > 0$ for all $a, b, c \in \mathcal{L}$.

cannot be accommodated with standard model averaging techniques (of, e.g. Debarsy and LeSage, 2020; Zhang and Yu, 2018).

The vector $\boldsymbol{\xi}$ scales the strength of incoming connectivity for individual units, giving weights to the neighbors indicated by $\boldsymbol{\Psi}$. It is not separately identified from the overall strength parameter λ, and is usually set in a way that guarantees a convenient domain for λ. As we will see below, this amounts to scaling the spectral radius of the adjacency matrix, $\rho(\boldsymbol{\Psi})$, to unity. With this goal in mind, there are two options with strong implications for the interpretation of spillover effects.

First, $\boldsymbol{\xi}$ can be set such that individual units are scaled differently. This is the case for the arguable standard choice, $\xi_i^{-1} = \sum_{j=1}^{n} [\boldsymbol{\Psi}]_{ij}$ for all i, which results in a row-stochastic connectivity matrix. Row-stochastic matrices yields spatial lags that are averages of neighboring values — unweighted for binary adjacency matrices, or weighted by the entries of $\boldsymbol{\Psi}$ for general ones. This transformation is convenient for interpretation, but does not retain symmetry in $\boldsymbol{\Psi}$ and imposes homogeneity on the overall connectivity of individual units (see Kelejian and Prucha, 2010; LeSage and Pace, 2014; Neumayer and Plümper, 2016, for further discussion).

Second, all elements of $\boldsymbol{\xi}$ can be set to some scalar, preserving potential symmetry of $\boldsymbol{\Psi}$ and retaining differences in connectivity between units. While there is no consensus on the exact scalar in the literature, $\xi_i^{-1} = \|\boldsymbol{\Psi}\|$, where $\|\cdot\|$ denotes some matrix norm, is a common choice. A notable example is $\min\{\|\boldsymbol{\Psi}\|_1, \|\boldsymbol{\Psi}\|_\infty\}$, proposed by Kelejian and Prucha (2010). As we will see below, the spectral radius of the adjacency matrix, $\rho(\boldsymbol{\Psi})$, is a sensible choice for general matrices. Since the spectral radius is an infimum of all norms, the domain of λ is generally more constrained than it needs to be in the earlier literature.

With this in mind, we are ready to re-express the connectivity function in Equation 4 in an even more verbose form, as

$$f(\lambda \mid \mathbf{W}) = f(\lambda \mid \boldsymbol{\xi}, g, \delta, d, \mathcal{L}). \tag{7}$$

This formulation makes obvious the many parameters and choices behind models that incorporate connectivity between units. The strong assumption that is fixing $\mathbf{W}$ is evident, and the motivation for more flexible specifications, which free up structural parameters, should be clear.

2.1.2 Connectivity strength

In models that include a connectivity filter, we want to ensure that it is (i) invertible, and (ii) stationary. Here, I provide conclusive results that guarantee these two properties for sufficiently general connectivity matrices. Specifically, we will assume the following form of the connectivity matrix.

Assumption 1. *>Vf7A TVIS `a` Ž WISf[hW/dV3^_ Sfdlj i [fZ $a_{ii} = 0$ Xad S^^iž*

This assumption is not prohibitive, and suffices to obtain the following result.

Theorem 1. *>Vf7I VWafVfZW[WWf[fk _ Sfdlj ł S` V α TVIS dV3^eLS′Sdŧ FZW $(\mathbf{I} - \alpha \mathbf{A})$ [e `a` Ž ef Yg′Sd Xad α $\in \left(\omega_{\min}^{-1}, \omega_{\max}^{-1} \right)$ ł i ZVdN/ω SdVfZW/dV3^V]YWhS′gVeaXAž*

BdaaXt This statement is true if $\alpha \cdot \omega_i \neq 1$ for all i, which we will show directly. For $\omega_i = 0$, this is trivially the case; we need to show it for all $\omega_i \neq 0$. Notice that trace($\mathbf{A}$) = 0, which implies that $\omega_{\min} < 0$ and $\omega_{\max} > 0$. In order to show our result, we have two requirements — for (1) positive eigenvalues it is enough to ensure that $\alpha < \omega_i^{-1}$, and for (2) negative ones that $\alpha > \omega_i^{-1}$. The result follows from knowing that $\omega_{\min}^{-1} < \alpha < \omega_{\max}^{-1}$. □

Non-singularity ensures the existence of the filter, and the continuity of its determinant, $|\mathbf{S}|$. Next, we are concerned with its stationarity, which can be understood as the notion that the (absolute) connectivity is decreasing with the order of neighbors — i.e. units are more connected to their direct neighbors than to their neighbors' neighbors. We can guarantee stationarity of the filter if its Neumann series is convergent, or equivalently, if its spectral radius of $\mathbf{S}$ is less than unity. This gives us the domain $\lambda \in \left(-\rho(\mathbf{W})^{-1}, \rho(\mathbf{W})^{-1} \right)$, which constrains our previous result, either from above or below.

For more concrete bounds, it helps to realize that $\mathbf{W}$ is either the null matrix, or is composed of null and irreducible submatrices, up to an isomorphism. Then Perron-Frobenius theorem gives us the following corollary.

Corollary 1. *FZWebWfdS^dSV[geaX $(\mathbf{I} - \alpha \mathbf{A})$ [e ′WdefZS` g` [fk Xad α $\in \left(-\omega_{\max}^{-1}, \omega_{\max}^{-1} \right)$ ž*

As can be seen, conclusive bounds for λ are determined by the spectral radius of $\mathbf{W}$, which coincides with the maximum real eigenvalue. This result is good news for the literature, where the usual domain of $\lambda \in (-1, 1)$ is supported by the standard choices for $\boldsymbol{\xi}$.

For large connectivity matrices, the prospect of computing its spectral radius may appear daunting, especially when that matrix is mutable. Direct methods for determining eigenvalues are computationally prohibitive, at a complexity of $\mathcal{O}(n^3)$. Iterative methods, such as

the Lanczos method or Arnoldi iteration, can provide remedy (Trefethen and Bau, 1997). These methods allow us to only compute the required largest eigenvalue, converge to an exact result at a general complexity of $\mathcal{O}(n^2)$, and particularly benefit from sparsity, which is a common feature of larger networks. Another option are approximations, such as matrix norms or ones based on the Gershgorin circle theorem. These only involve comparatively trivial computations, and usually produces bounds that are tighter than the ones implied by the spectral radius.

2.2   Interpretation

The interpretation of spillover effects in the general spatial econometric model is not entirely straightforward, and depends on the spatial lags involved. Firstly, the lag of the error term does not directly convey the presence of spillovers, but implies correlation across observations (see Barrios et al., 2012, for a comparison with clustering). Next, the spatially lagged explanatory variables represent 'local' spillovers from the weighted characteristics of direct neighbors (see Halleck Vega and Elhorst, 2015, for further discussion). Lastly, the endogenous lag of the response variable represents 'global' spillovers across all neighbors. It is arguably the most widely used, but also the most controversial term (see Angrist, 2014; Corrado and Fingleton, 2012), and is best understood as the reflection of an equilibrium response.

In the standard spatial econometric model, the free parameters can only be interpreted conditional on the connectivity structure. This circumstance is a common source of errors and misconceptions (LeSage and Pace, 2014; Kuschnig, 2022), and holds similarly for the more complex models considered in this paper. Most of these issues stem from a misinterpretation of the partial effects. For the model in Equation 1, the partial effects of a variable j are given by

$$
\frac{\partial \mathbf{y}}{\partial \mathbf{x}_k} = \mathbf{S}(\lambda \mid \mathbf{W})^{-1} \begin{pmatrix} \beta_k & w_{12}\theta_j & \dots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \dots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \dots & \beta_k \end{pmatrix}.
\tag{8}
$$

This matrix of partial effects is not straightforward to interpret, and summary measures are used instead. An obvious example is the average partial effect, for which we divide the sum of all elements in Equation 8 by n. LeSage and Pace (2009) consider two summary measures that offer insights into the spatial structure — the average direct effect (considering only

diagonal elements), and the average indirect effect (considering only off-diagonal elements).

3   Priors

In this section, I present suitable priors for the model in Equations 2 and 3, including options for more flexible connectivity specifications (following Equation 7). For the sake of completeness, we start with the nested linear model in Equation 2.

For the linear model, there is a variety of established priors to choose from, including the standard Normal Inverse-Gamma prior or a number of shrinkage priors (e.g. Bhattacharya et al., 2015; Griffin and Brown, 2010; Park and Casella, 2008, etc.). We will consider the Horseshoe prior by Carvalho et al. (2010) for its proven performance and efficient sampling (see Makalic and Schmidt, 2015), and an inverse Gamma prior for the variance. That is

$$\beta_i \mid \iota_i \sim \mathrm{N}(0, \iota_i), \qquad \iota_i \mid \zeta \sim \mathrm{C}^+(0, \zeta), \qquad \zeta \mid s \sim \mathrm{C}^+(0, \varsigma), \tag{9}$$
$$\sigma^2 \sim \mathrm{G}^{-1}(a, b),$$

where ι_i are coefficient-specific hyperparameters with common scale t, and $\mathrm{C}^+(0, a)$ denotes the density of the half-Cauchy distribution with scale a, given by

$$\mathrm{C}^+(x \mid 0, a) = \begin{cases} \frac{2}{\pi a}\left(1 + \frac{x^2}{a^2}\right)^{-1}, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

The Horseshoe prior offers flexible shrinkage, accommodating sparsity via sharp spikes at the origin without distorting estimates, thanks to its heavy tails. The parameters facilitate global regularization, but also allow for local shrinkage of individual coefficients. As a result, the Horseshoe is a standard choice in the literature (also see Datta and Ghosh, 2013; van der Pas et al., 2014, 2017). This allows us to focus on the connectivity going forward.

3.1   Hierarchical shrinkage for connectivity strength

In the wider spatial econometric literature, the overall connectivity strength λ is the only free parameter, and plays a central role when analyzing connectivity. With a standardized connectivity matrix, its domain is $(-1, 1)$, and it is most commonly assigned a Uniform prior, generalized by the Beta prior (first proposed by LeSage and Parent, 2007)

$$p(\bar{\lambda}) \sim \mathrm{Be}(1 + \tau, 1 + \tau),$$

where $\tau \geq 0$, and we use $\bar{\lambda} = (\lambda + 1)/2$, scaled to live on $(0, 1)$, for simplicity. $\mathrm{Be}(a, b)$ denotes the density of a Beta distribution with shapes a and b, i.e.

$$\mathrm{Be}(x \mid a, b) = \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{Beta}(a, b)}.$$

For $\tau = 0$, the prior is uniform over all values. With this in mind, the prior parameter τ can be understood as excess support for the origin — but, as we will see, this interpretation is misleadingly narrow.

Beyond the Uniform prior, we sometimes encounter $\tau = 0.01$ in the literature, which results in a fairly flat prior that places slightly higher weight at the origin. On one hand, this reflects a clear orientation on flat priors that are uninformative with respect to certain values. On the other hand, however, this is driven by necessity due to the Beta distribution's undesirable properties. These undesirable properties are moderate peaks in density, and excessive drop-offs towards the tails, as illustrated in Figure 1. There, we can see three Beta densities, placing increasing mass at the origin. Even for $\tau = 100$, the peak remains moderate, while the density at the tails (e.g. at 0.9) becomes miniscule, and the credible support of the prior incredibly narrow.

With any prior distribution, we want to express the prior information that is available to us without distorting insights that we can obtain from the data. For our prior for λ, we thus want to incorporate information and avoid potential distortions. First, consider the often implicit, and sometimes explicit preference for parsimony, i.e. $\lambda = 0$. Without decent support for non-zero values, it appears reasonable to skip the superfluous complexity of connectivity. This potentially special role of zero highlights the distorting information induced by the Uniform prior. While this flat prior indicates no preference for any values, it implicitly prefers large values, essentially imposing connectivity on the model. For instance, we have $p(|\lambda| > 0.1) = 9 \cdot p(|\lambda| \leq 0.1)$, or visually, ▉━━━━━━.

This motivates our departing point, which is the the following mixture prior

$$p(\bar{\lambda}) \sim \begin{cases} \mathrm{Be}(1 + \tau_0, 1 + \tau_0), & \text{if } \gamma = 1, \\ \mathrm{Be}(1 + \tau_1, 1 + \tau_1), & \text{if } \gamma = 0, \end{cases}$$

where $\tau_0 \ll \tau_1$ are shape parameters, and γ is an indicator. This prior essentially represents a variable selection procedure for λ. For $\gamma = 0$, the sharp spike at zero that is induced by τ_1 leads to a collapse to the linear model. For $\gamma = 1$, we have a comparatively flat prior for λ, mirroring standard setups. This small adaptation to a spike-and-slab prior takes us in the right direction conceptually, but arguably remains too rigid (except, perhaps, for panels
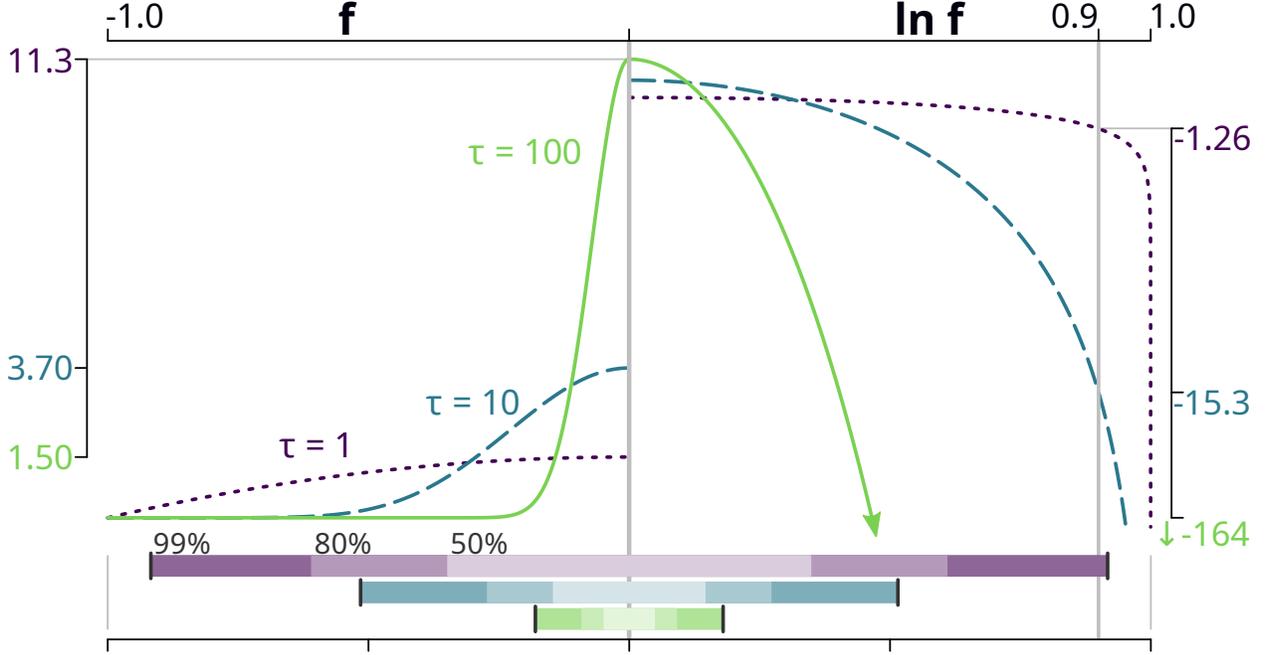
Figure 1: Density (left), log-density (right), and 99%, 80%, and 50% credible intervals (bottom) for λ with a Beta prior, i.e. $p(\bar{\lambda}) \sim \text{Be}(1 + \tau, 1 + \tau)$, with $\tau \in \{1, 10, 100\}$. More informative priors (with increasing τ) lead to narrow credible intervals, with values in the tail (e.g. 0.9, compare the right panel) receiving infinitesimal prior support.

with time-specific connectivity). Next, we consider further sources of prior information, and develop a practical prior that can accommodate our prior convictions.

The second source of prior information we want to reflect concerns the model specification — in particular, the boundaries of λ. Parameters that lie at their boundary commonly cause issues for statistical models (Chernoff, 1954; Self and Liang, 1987; Chen and Liang, 2010), and λ is no exception. Numerical instability from the filter approaching singularity is an obvious example, but not the most damning. Instead, a major issue are pathological solutions[3] that arise, not from the phenomenon under investigation, but the peculiar structure of the model (see, for example, Angrist, 2014; Halleck Vega and Elhorst, 2015). In the face of these technical and theoretical caveats, it is sensible to regularize λ and limit the support at the boundaries by imposing a suitable prior.

I propose the following continuous mixture of Beta distributions as a prior, which can coalesce the above sources of prior information in a straightforward way.

$$p(\bar{\lambda}) \sim \text{Be}(1 + \tau, 1 + \tau), \qquad p(\tau) \sim \text{Ga}(a, b), \tag{10}$$

---

[3]One example for a model considered here, is the perfect fit from $\lambda \to -1$, $\delta \to 0$, and $\alpha = \sum_{i=1}^{n} y_i$.

where the mixing density is a Gamma distribution with shape a and rate b, which has proven to be useful for this purpose in similar settings (see Park and Casella, 2008; Griffin and Brown, 2010). This mixture affords us considerable flexibility for prior elicitation, allowing us to act upon the prior information discussed above.
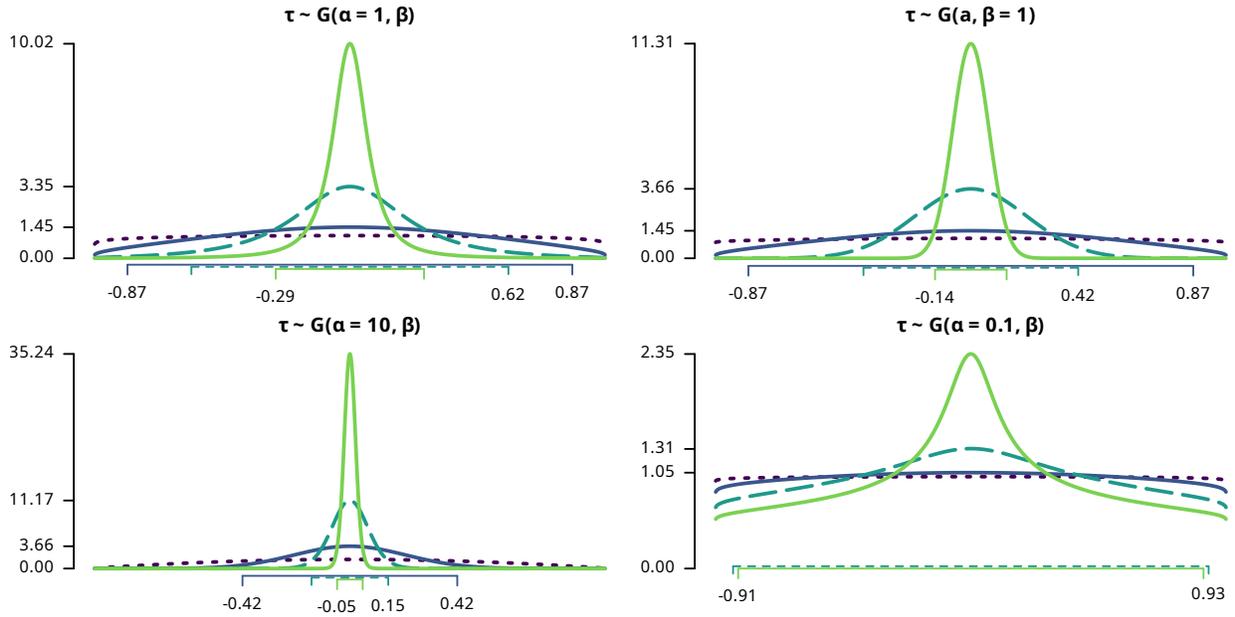
Figure 2: Density (left), log-density (right), and 99%, 80%, and 50% credible intervals (bottom) with a Beta-Gamma prior.

The Beta-Gamma mixture can accommodate a wide range of shapes, as exemplified in Figure 2. We can place considerable mass at the origin, while also accommodating values in the tails. If we constrain the prior to the special case of an Exponential ($a = 1$, top-left) and orient ourselves on Figure 1, we can clearly see that the mixture prior yields more pronounced peaks with wider credible support and without excessive drop-offs. By varying both parameters, we are able to flexibly induce fine-tuned priors, essentially without overhead.

The proposed Beta-Gamma shrinkage prior changes the prior specification from an issue of choosing specific values, to one concerned with parsimony and regularization. Increased weight at the origin means that the prior does not induce spillover effects per se, while the tail behavior allows us to provide regularization without limiting support to narrow regions a priori. The result is a flexible prior that can better express many prior convictions. Nonetheless, there may be reservations to even weakly informative priors.[4] While they may

---

[4]Ignore, for this example, that the standard flat priors are heavily informative in terms of size.

not always be a necessity, it is important to keep in mind that we introduce this structure to free up the previously implicit and infinitely informative priors that are fixed structural parameters.

3.2   Priors for the connectivity structure

Equation 7 highlights how current models feature the strong assumption that connectivity structures are known; only its strength is uncertain. Here, we focus on freeing up and modelling explicitly two different structural parameters — (1) the number of neighbors k, and (2) the speed of distance decay δ. For each, I present suitable prior distributions that allow for more flexible and, thus, credible models.

k-nearest neighbor structures

The first candidate for more explicit treatment is the number of neighbors in the k-nearest neighbor specification in Equation 5. Limited variations to the parameter are commonly considered for robustness checks, and have previously been addressed within a model averaging framework (see Lesage and Fischer, 2008; Debarsy and LeSage, 2020; Zhang and Yu, 2018, for instance). While the discrete parameter lends itself to model averaging, we will take it one step further and treat k in a fully Bayesian way. For this, we can consider it as the result of N trials that determine whether any two units are neighbors or not. Such a trial can be modelled with a Beta-binomial distribution, i.e.

$$p(k) \sim \mathrm{BB}(N, a, b),$$

where N is the number of trials (i.e. potential neighbors), a and b describe the probability of success, and $\mathrm{BB}(N, a, b)$ denotes the density of a Beta-binomial distribution, i.e.

$$\mathrm{BB}(x \,|\, N, a, b) = \binom{N}{x} \frac{\mathrm{Beta}(a + x, b + N - x)}{\mathrm{Beta}(a, b)}.$$

With this prior, there is no need to constrain our model to a selection of values. We can open the full model space by setting $N = n - 1$, retain sparsity in connections by choosing $a < b$ appropriately, and allow for efficient estimation using Markov chain Monte Carlo methods.

In the literature, there is a strong preference for few neighbors, both expressed in empirical work and for theoretical reasons. In Figure 3, we can see that the Beta-Binomial prior allows for nuanced priors that reflect this preference. However, it mirrors the underlying Beta distribution, in that it experiences excessive drop-off in the tails. If this aspect is
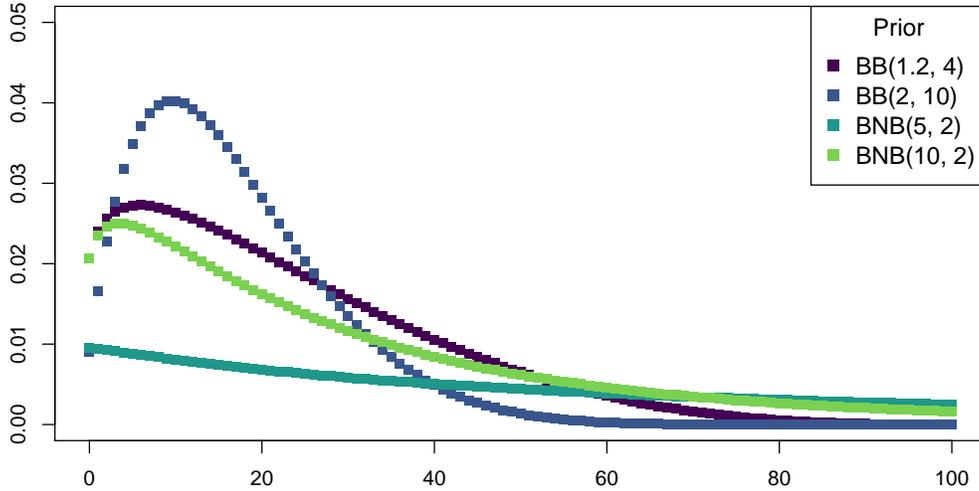
Figure 3: Visualization of different prior setups for *k*, with $n = 100$.

considered restrictive, a Beta-negative-Binomial prior (visualized) can present an even less informative alternative.

Distance decay structures

Equation 6 introduces another common type of connectivity structure, which is based on distance-decay and a decay parameter δ. Earlier works that are limited to local lags show that there is impactful uncertainty around this parameter (Halleck Vega and Elhorst, 2015; Kuschnig, 2022). Standard model averaging approaches provide no remedy due to the continuous nature of δ. With our fully Bayesian approach, we merely need a sensible prior for the parameter. A useful option is

$$\delta \sim \mathrm{Ga}^{-1}(a, b). \tag{11}$$

The inverse-Gamma distribution is flexible enough to accommodate general prior conceptions. For our parameter, an important benefit is that it avoids placing weight at and near zero values, where every unit is equally connected to every other unit. If these small values at the left tail are a problem, a log-Normal prior can be a useful alternative. When strong and explicit prior information is available, the Weibull distribution can be another useful alternative. For a visualization of selected priors, see Figure 4.

Regarding specific values for δ, the literature is not particularly informative. While suitable values depend on the distances involved, $\delta = 1$ can serve as an anchor for prior elicitation. Below it, i.e. for $\delta \ll 1$, connections between neighbors are weighted more
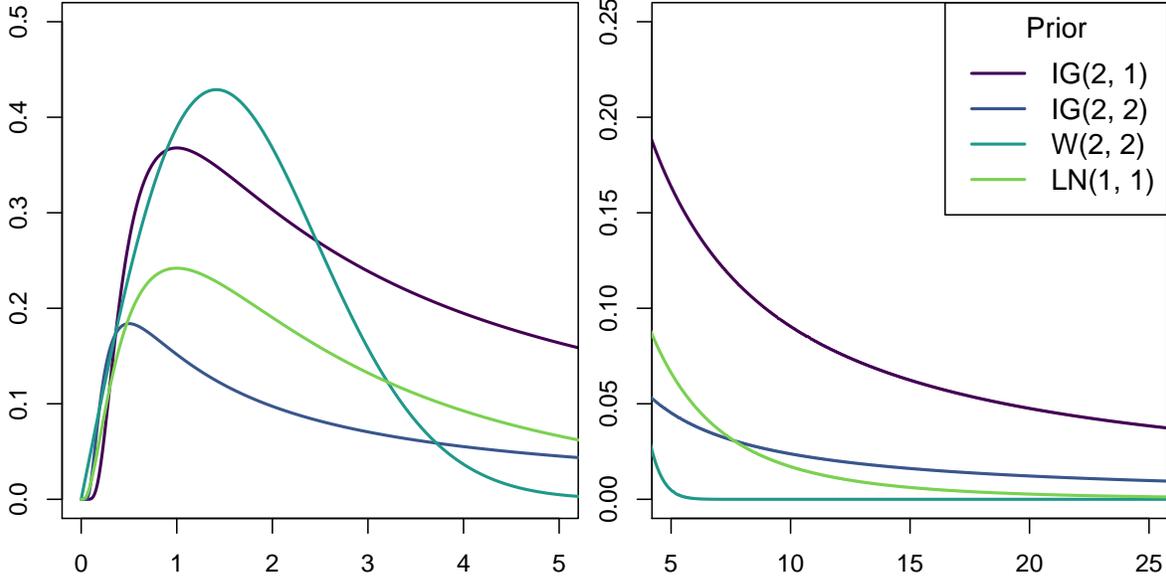
Figure 4: Visualization of different prior setups for δ. The left panel highlights behaviour near the focal value of $\delta = 1$; the right panel shows the tail behavior.

equally, with less regard to the distance. For $\delta \gg 1$, by contrast, only the closest neighbors retain relevance as connectivity levels off faster than the distance.

4 Estimation

We consider the model in Equations 2 and 3, with the connectivity given by $f(\lambda, \delta \mid \cdot)$, and use the priors outlined in Equations 9, 10, and 11. In this section, we describe a Markov chain Monte Carlo (MCMC) approach to obtain full posteriors of this setup.

4.1 Sampling

First, note that we can readily obtain posterior draws of $\left(\beta, \sigma^2\right)$ conditional on (λ, δ, τ) using the approach by Makalic and Schmidt (2015). Next, we draw from the conditional posterior of λ, and then τ. Finally, we draw from the conditional posterior δ, and repeat — giving us the procedure in Figure 5.

For the first step, we can rely on standard techniques, such as Gibbs sampling, by conditioning on the connectivity parameters. In the second and third steps, the conditional posteriors of λ, τ, and δ have no well-known form, and we must use another approach.

0. Set starting values for $\lambda, \delta, \tau, \sigma^2$.

1. Draw from the conditional posteriors of the nested linear model, i.e.

   (a) $p\left(\boldsymbol{\beta}|\mathbf{y}, \lambda, \delta, \tau, \sigma^2\right)$,

   (b) $p\left(\sigma^2|\mathbf{y}, \lambda, \delta, \tau, \boldsymbol{\beta}\right)$.

2. (a) Draw from $p\left(\lambda|\mathbf{y}, \boldsymbol{\beta}, \sigma^2, \delta, \tau\right)$, and

   (b) $p\left(\tau|\mathbf{y}, \lambda, \cdot\right)$.

3. Draw from $p\left(\delta|\mathbf{y}, \boldsymbol{\beta}, \sigma^2, \delta\right)$.

4. Go to the first step until enough draws are obtained.

Figure 5: Stylized algorithm for sampling from the model.

For λ, we can use a Metropolis-Hastings step to draw from its conditional posterior,

$$p(\lambda|\mathbf{y}, \tau, \cdot) \propto |\mathbf{S}(\lambda, \delta)| \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right\} p(\lambda|\tau).$$

The conditional posterior of τ can be expressed as

$$p(\tau|\mathbf{y}, \lambda, \cdot) \propto \lambda^\tau (1-\lambda)^\tau \tau^{a-1} \exp^{-\tau b},$$
$$\propto \tau^{a-1} \exp^{-\tau\left[b - \log\left(\lambda - \lambda^2\right)\right]},$$

which is the kernel of a Gamma density, which we can directly draw from using a Gibbs step. This means that our hierarchical prior setup for λ imposes essentially no overhead over conventional specifications.

Lastly, another Metropolis-Hastings step allows us to draw from

$$p(\delta|\mathbf{y}, \cdot) \propto |\mathbf{S}(\lambda, \delta)| \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{S}(\lambda, \delta)\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right\} p(\delta).$$

With the exception of τ, these sampling steps are well-known, and all of them are conceptually straightforward. However, they pose one major computational challenge — that is, the determinant of the $n \times n$ Jacobian matrix $\mathbf{S}(\lambda, \delta)$.

4.2 Evaluating the Jacobian determinant

The likelihood, and hence the posterior, of our model involves a Jacobian determinant, which poses a central computational constraint for estimation (Bivand et al., 2013). In standard

models, we can use a spectral decomposition of the fixed connectivity matrix $\mathbf{W}$ to compute the determinant with the eigenvalue method, using

$$\ln |\mathbf{I} - \lambda \mathbf{W}| = \sum_{i=1}^{n} \ln \left(1 - \lambda \omega_i \right).$$

There are other approaches for large matrices, e.g. based on the lower-upper decomposition, spline approximations, or algebraic results that make use of special connectivity structures (see Bivand et al., 2013). However, all of these approaches rely on the connectivity structure in $\mathbf{W}$ being fixed, and would thus present a potentially insurmountable computational challenge for more flexible models.

In order to still allow for rapid estimation using MCMC, we introduce the following Gaussian process approximation

$$|\mathbf{S}(\lambda, \delta)| \approx \mathrm{GP} \left(\mu(\lambda, \delta), \Sigma(\lambda, \delta) \right),$$

This allows us to approximate the Jacobian determinant for credible values of (λ, δ) with high accuracy (cf. the Supplementary Material). Essentially, we compute the eigenvalues for a grid of δ values, use those to determine $|\mathbf{S}(\lambda, \delta)|$ using the eigenvalue method, and fit these training samples using Gaussian process regression. This approach provides a quantification of uncertainty, and allows for retraining if the sampler moves to values far from the grid. For our approach, we rely on a constant mean, μ, and a Gaussian kernel for Σ, but other options are available. Notably, Gaussian processes are widespread in the field of spatial statistics, and have a parallel in spline regression, which can be used to approximate one-dimensional Jacobian determinants.

5   Illustration

To illustrate the hierarchical model developed in this paper, we will revisit the issue of deforestation spillovers from croplands in Mato Grosso, Brazil (following Kuschnig et al., 2021). We use the original panel dataset of 141 municipalities from 2006–2017, and consider an autoregressive model with inverse-distance decay based connectivity matrices. The model is given by

$$\mathbf{y}_t = \lambda \mathbf{\Psi}(\delta) \mathbf{y}_t + \mathbf{X}_{t-1} \boldsymbol{\beta} + \boldsymbol{\mu} + \boldsymbol{\psi}_t + \boldsymbol{\varepsilon}_t.$$

The dependent variable is the change of forest per area, and the regressors are the shares of forest, pasture, and cropland area, population density, cattle density (per pasture), soy
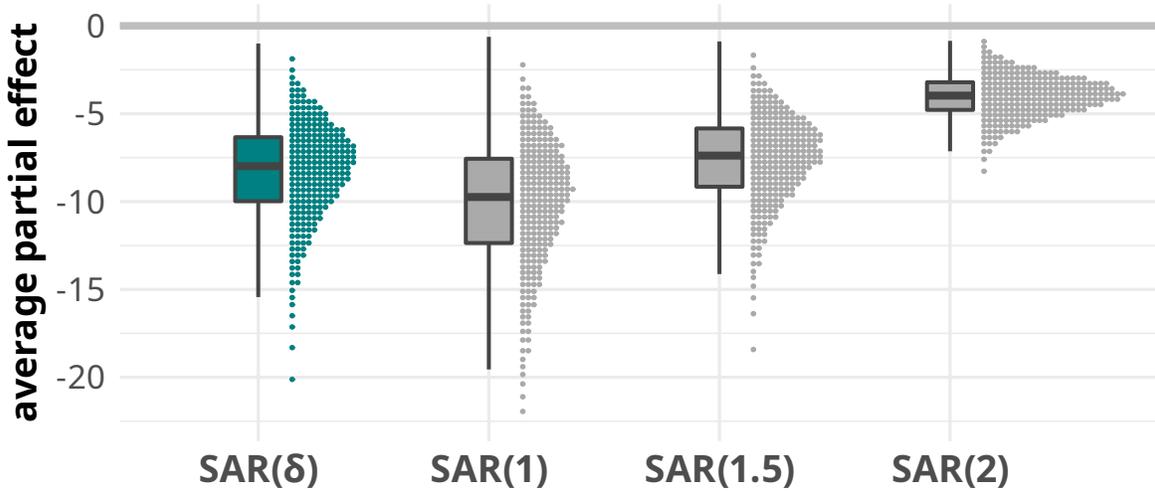
Figure 6: Visualization of the average partial effect of croplands on deforestation.

yields in Brazilian real per harvested area, and an indicator for the incidence of particularly dry months. We will focus on the partial effect of croplands, which is given by

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}_{\text{crops}}} = \mathbf{S}(\lambda, \delta)^{-1} \left(\mathbf{I}\beta_{\text{crops}} \right).$$

In Figure 6, we can see the average partial effect of croplands (direct and indirect) of our model of choice, where δ, the speed of distance-decay, is modelled explicitly. We can see that a naive specification that fixes $\delta = 2$ underestimates both the impact of croplands, and the uncertainty surrounding it. For $\delta = 1$, both the impact and the uncertainty is overestimated instead. If the parameter is fixed close to its posterior mean, i.e. $\delta = 1.5$, as one might do when adopting an empirical Bayes approach, the uncertainty is underestimated.

6   Conclusion

In this paper, I introduced a comprehensive framework for modelling connectivity between units of observations, and the resulting spillover effects. I used a Bayesian hierarchical approach that is readily extensible, allows for regularization that helps in freeing up parameters, and natively conveys uncertainty. For this purpose, I decomposed the standard spatial econometric connectivity (or weights) matrix, and derived suitable bounds to guarantee non-singularity and stationarity. I introduced the flexible Beta-Gamma mixture prior for the parameter of overall connectivity strength in autoregressive models, and discussed suitable priors for parameters that determine the connectivity structure. I presented an efficient

sampling approach that is readily extensible and allows for full posterior inference. To allow for efficient sampling with a number of free connectivity parameters, I proposed a Gaussian process approximation for the Jacobian determinant in autoregressive models. The result is a flexible and extensible framework for connectivity models that works in general settings.

References

Daron Acemoglu, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. The network origins of aggregate fluctuations. *7Ua`a_ Vfd[US*, 80(5):1977–2016, 2012. ISSN 0012-9682. doi:10.3982/ECTA9623.

Daron Acemoglu, Ufuk Akcigit, and William R. Kerr. Innovation network. *BdaUWW[`Ye aX fZW @Sf[a`S^ 3USVW_k aX EU[W`UWe*, 113(41):11483–11488, 2016. ISSN 0027-8424. doi:10.1073/pnas.1613559113.

Attila Ambrus, Markus Mobius, and Adam Szeidl. Consumption risk-sharing in social networks. *3_Wd[US` 7Ua`a_[UDWh[Wi*, 104(1):149–82, 2014. ISSN 0002-8282. doi:10.1257/aer.104.1.149.

Joshua D. Angrist. The perils of peer effects. *>STagd 7Ua`a_[Ue*, 30:98–108, 2014. ISSN 0927-5371. doi:10.1016/j.labeco.2014.05.008.

Enghin Atalay, Ali Hortaçsu, James Roberts, and Chad Syverson. Network structure of production. *BdaUWW[`Ye aXfZW@Sf[a`S^3USVW_k aXEU[W`UWe*, 108(13):5199–5202, 2011. ISSN 0027-8424. doi:10.1073/pnas.1015564108.

Enghin Atalay, Ali Hortaçsu, and Chad Syverson. Vertical integration and input flows. *3_Wd[US` 7Ua`a_[UDWh[Wi*, 104(4):1120–48, 2014. ISSN 0002-8282. doi:10.1257/aer.104.4.1120.

Thomas Barrios, Rebecca Diamond, Guido W. Imbens, and Michal Kolesár. Clustering, spatial correlations, and randomization inference. *<agd`S^aXfZW3_Wd[US` EfSf[ef[US^3eeaU[Sf[a`*, 107 (498):578–591, 2012. ISSN 0162-1459. doi:10.1080/01621459.2012.682524.

Lori A. Beaman. Social networks and the dynamics of labour market outcomes: evidence from refugees resettled in the U.S. *DWh[Wi aX7Ua`a_[UEfgV[We*, 79(1):128–161, 2012. ISSN 0034-6527. doi:10.1093/restud/rdr017.

Anirban Bhattacharya, Debdeep Pati, Natesh Pillai, and David Dunson. Dirichlet-Laplace priors for optimal shrinkage. *<agd`S^aXfZW3_Wd[US` EfSf[ef[US^3eeaU[Sf[a`*, 110(512):1479–1490, 2015. doi:10.1080/01621459.2014.960967.

Roger Bivand, Jan Hauke, and Tomasz Kossowski. Computing the Jacobian in Gaussian spatial autoregressive models: an illustrated comparison of available methods. *9VaYdSbZ[US^3` S^ke[e*, 45 (2):150–179, 2013. doi:10.1111/gean.12008.

Nicholas Bloom, Mark Schankerman, and John Van Reenen. Identifying technology spillovers and product market rivalry. *7Ua`a_ Wfd[US*, 81(4):1347–1393, 2013. ISSN 0012-9682. doi:10.3982/ECTA9466.

Simon Board and Moritz Meyer-ter Vehn. Learning dynamics in social networks. *7Ua`a_ Wfd[US*, 89 (6):2601–2635, 2021. ISSN 1468-0262. doi:10.3982/ECTA18659.

Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. The horseshoe estimator for sparse signals. *4[a_ Wfd[]S*, 97(2):465–480, 2010. doi:10.1093/biomet/asq017.

Yong Chen and Kung-Yee Liang. On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. *4[a_ Wfd[]S*, 97(3):603–620, 9 2010. ISSN 0006-3444. doi:10.1093/biomet/asq031.

Herman Chernoff. On the distribution of the likelihood ratio. *FZW3` `S^e aX? SfZW Sf[US^EfSf[ef[Ue*, pages 573–578, 1954. doi:10.1214/aoms/1177728725.

Eric Chyn and Lawrence F. Katz. Neighborhoods matter: assessing the evidence for place effects. *<agd ` S^aX7Ua`a_ [UBWdebWUf[hWe*, 35(4):197–222, 2021. ISSN 0895-3309. doi:10.1257/jep.35.4.197.

Luisa Corrado and Bernard Fingleton. Where is the economics in spatial econometrics? *<agd S^aX DW[a` S^EU[WUW*, 52(2):210–239, 2012. doi:10.1111/j.1467-9787.2011.00726.x.

Jesús Crespo Cuaresma, Gernot Doppelhofer, Martin Feldkircher, and Florian Huber. Spillovers from US monetary policy: evidence from a time varying parameter global vector autoregressive model. *<agd S^aXfZWDakS^EfSf[ef[US^EaU[Wfk, EWd[We 3 /EfSf[ef[Ue [EaU[Wfkf*, 182(3):831–861, 2019. doi:10.1111/rssa.12439.

Jyotishka Datta and Jayanta. K. Ghosh. Asymptotic properties of Bayes risk for the Horseshoe prior. *4SkWe[S` 3` S^ke[e*, 8(1):111–132, 2013. ISSN 1936-0975. doi:10.1214/13-BA805.

Nicolas Debarsy and James P. LeSage. Bayesian model averaging for spatial autoregressive models based on convex combinations of different types of connectivity matrices. *<agd S^aX4ge[` We ˜ 7Ua`a_ [UEfSf[ef[Ue*, pages 1–12, 2020. doi:10.1080/07350015.2020.1840993.

Lorenzo Ductor, Marcel Fafchamps, Sanjeev Goyal, and Marco J. van der Leij. Social networks and research output. *DWh[Wi aX7Ua`a_ [Ue S` V EfSf[ef[Ue*, 96(5):936–948, 2014. ISSN 0034-6535. doi:10.1162/REST_a_00430.

Michael Gofman. Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *ᐊagd′ S^ aX 8[′ S′ U[S^ 7Ua′ a_ [Ue*, 124(1):113–146, 2017. ISSN 0304-405X. doi:10.1016/j.jfineco.2016.12.009.

Paul Goldsmith-Pinkham and Guido W. Imbens. Social networks and the identification of peer effects. *ᐊagd′ S^ aX 4ge[′ Wde ˘ 7Ua′ a_ [U EfSf[ef[Ue*, 31(3):253–264, 2013. doi:10.1080/07350015.2013.801251.

Jim E. Griffin and Philip J. Brown. Inference with Normal-Gamma prior distributions in regression problems. *4SkW[S′ 3′ S′ke[e*, 5(1):171–188, 2010. doi:10.1214/10-BA507.

Solmaria Halleck Vega and J. Paul Elhorst. The SLX model. *ᐊagd′ S^aXDW[′a′S^EU[WU[/*55(3): 339–363, 2015. doi:10.1111/jors.12188.

Lena Hensvik and Oskar Nordström Skans. Social networks, employee selection, and labor market outcomes. *ᐊagd′ S^aX>STad 7Ua′a_ [Ue*, 34(4):825–867, 2016. doi:10.1086/686253.

Chih-Sheng Hsieh and Lung Fei Lee. A social interactions model with endogenous friendship formation and selectivity. *ᐊagd′ S^aX3bb[′W 7Ua′a_ Wfd[Ue*, 31(2):301–319, 2016. doi:10.1002/jae.2426.

Yannis M. Ioannides and Linda Datcher Loury. Job information networks, neighborhood effects, and inequality. *ᐊagd′ S^ aX 7Ua′ a_ [U >[fWdSfgdW/* 42(4):1056–1093, 2004. ISSN 0022-0515. doi:10.1257/0022051043004595.

Ida Johnsson and Hyungsik Roger Moon. Estimation of peer effects in endogenous social networks: control function approach. *DWh[W aX7Ua′a_ [UeS′V EfSf[ef[Ue*, 103(2):328–345, 5 2021. ISSN 0034-6535. doi:10.1162/rest_a_00870.

Harry H. Kelejian and Ingmar R. Prucha. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *ᐊagd′ S^aX7Ua′a_ Wfd[Ue*, 157(1):53–67, 7 2010. ISSN 0304-4076. doi:10.1016/j.jeconom.2009.10.025.

Michael D. König, Xiaodong Liu, and Yves Zenou. R&D networks: theory, empirics, and policy implications. *DWh[W aX 7Ua′a_ [Ue S′ V EfSf[ef[Ue*, 101(3):476–491, 2019. ISSN 0034-6535. doi:10.1162/rest_a_00762.

Rachel E. Kranton and Deborah F. Minehart. A theory of buyer-seller networks. *3_ Wd[US′ 7Ua′a_ [U DWh[W*, 91(3):485–508, 2001. ISSN 0002-8282. doi:10.1257/aer.91.3.485.

Nikolas Kuschnig. Bayesian spatial econometrics: a software architecture. *ᐊagd′ S^aXEbSf[S^7Ua′aŽ _ Wfd[Ue*, 3(1):6–25, 2022. ISSN 2662-298X. doi:10.1007/s43071-022-00023-w.

Nikolas Kuschnig, Jesús Crespo Cuaresma, Tamás Krisztin, and Stefan Giljum. Spillover effects in agriculture drive deforestation in Mato Grosso, Brazil. *EU[Wf[UDWadfe*, 11(1):1–9, 2021. doi:10.1038/s41598-021-00861-y.

Clifford Lam and Pedro C. L. Souza. Estimation and selection of spatial weight matrix in a spatial lag model. *<agd S^ aX 4ge[Vde ˘ 7Ua`a_ [U EfSf[ef[Ue*, 38(3):693–710, 2020. doi:10.1080/07350015.2019.1569526.

James LeSage and Robert Kelley Pace. *;`fdaVgUf[a` fa ebSf[S^Wia`a_ Wff[Ue*. Chapman and Hall/CRC, 2009. doi:10.1201/9781420064254.

James P. Lesage and Manfred M. Fischer. Spatial growth regressions: model specification, estimation and interpretation. *EbSf[S^ 7Ua`a_ [U 3`S^ke[e*, 3(3):275–304, 2008. doi:10.1080/17421770802353758.

James P. LeSage and R. Kelley Pace. The biggest myth in spatial econometrics. *7Ua`a_ Wff[Ue*, 2(4): 217–249, 2014. doi:10.3390/econometrics2040217.

James P. LeSage and Olivier Parent. Bayesian model averaging for spatial econometric models. *9WaYdSbZ[US^3`S^ke[e*, 39(3):241–267, 2007. doi:10.1111/j.1538-4632.2007.00703.x.

Xu Lin. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *<agd S^aX>STad7Ua`a_ [Ue*, 28(4):825–860, 2010. doi:10.1086/653506.

Enes Makalic and Daniel F. Schmidt. A simple sampler for the horseshoe estimator. *;777 E[Y S^ BdaUWe[Y >Wff Wde*, 23(1):179–182, 2015. doi:10.1109/LSP.2015.2503725.

Angelo Mele. Does school desegregation promote diverse interactions? an equilibrium model of segregation within schools. *3_ Wd[US` 7Ua`a_ [U<agd S^, 7Ua`a_ [UBa^[Uk*, 12(2):228–57, 2020. ISSN 1945-7731. doi:10.1257/pol.20170604.

Kaivan Munshi. Networks in the modern economy: Mexican migrants in the U.S. labor market. *CgSdfWd k <agd S^ aX 7Ua`a_ [Ue*, 118(2):549–599, 2003. ISSN 0033-5533. doi:10.1162/003355303321675455.

Eric Neumayer and Thomas Plümper. W. *Ba^[f[US^EU[WVUWDWWSdUZ S`V ? Wf7aVe*, 4(1):175–193, 2016. doi:10.1017/psrm.2014.40.

Mark E. J. Newman. The structure of scientific collaboration networks. *BdaUWW[`Ye aXfZW@Sf[a` S^ 3USVW_ k aXEU[WUWe*, 98(2):404–409, 2001. ISSN 0027-8424. doi:10.1073/pnas.98.2.404.

Trevor Park and George Casella. The Bayesian lasso. *<agd S^aXfZW3_ Wd[US` EfSf[ef[US^3 eeaU[Sf[a`*, 103(482):681–686, 2008. doi:10.1198/016214508000000337.

Xi Qu and Lung-fei Lee. Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *◁agd′ S^aX7Ua`a_ Vfd[Ue*, 184(2):209–232, 2015. ISSN 0304-4076. doi:10.1016/j.jeconom.2014.08.008.

Andrew K. Rose. Do we really know that the WTO increases trade? *3_ Vd[US` 7Ua`a_ [UDwm[WV*, 94 (1):98–114, 2004. ISSN 0002-8282. doi:10.1257/000282804322970724.

Steven G. Self and Kung-Yee Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *◁agd′ S^aXfZW3_ Vd[US` EfSf[ef[US^3æaU[Sf[a`*, 82(398):605–610, 6 1987. ISSN 0162-1459. doi:10.1080/01621459.1987.10478472.

Lloyd N. Trefethen and David Bau. *@g_ Vd[US^ ′f̂ VgSd S′YVWdS.* SIAM, 1997. doi:10.1137/1.9780898719574.

S. L. van der Pas, B. J. K. Kleijn, and A. W. van der Vaart. The horseshoe estimator: posterior concentration around nearly black vectors. *7′WVfda` [U◁agd′ S^aXEfSf[ef[Ue*, 8(2):2585–2618, 2014. ISSN 1935-7524. doi:10.1214/14-EJS962.

Stéphanie van der Pas, Botond Szabó, and Aad van der Vaart. Uncertainty quantification for the Horseshoe (with discussion). *4SkVWf[S` 3` S′ke[e*, 12(4):1221–1274, 2017. ISSN 1936-0975. doi:10.1214/17-BA1065.

Ben Weidmann and David J. Deming. Team players: how social skills improve team performance. *7Ua`a_ Vfd[US*, 89(6):2637–2657, 2021. ISSN 1468-0262. doi:10.3982/ECTA18461.

Xinyu Zhang and Jihai Yu. Spatial weights matrix selection and model averaging for spatial autoregressive models. *◁agd′ S^aX7Ua`a_ Vfd[Ue*, 203(1):1–18, 2018. doi:10.1016/j.jeconom.2017.05.021.

A   Supplementary information

Spillover effects

The general spatial model gives rise to three spatial effects that are visualized in Figure A1. There may be local spillover effects, affecting direct neighbors, which arise from the lagged explanatories, $\mathbf{WX}\boldsymbol{\theta}$. The autoregressive term, $\lambda\mathbf{Wy}$, allows the responses of all units to be influenced by the response's of their neighbors. The observed equilibrium results from these global spillover effects, which spread across all connected units. Lastly, the error term may exhibit spatial autocorrelation due to $\rho\mathbf{We}$.
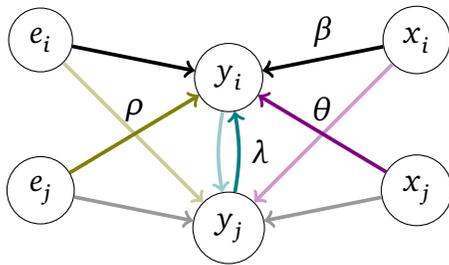

Figure A1: Illustration of a unit i receiving direct effects from itself, and indirect effects from other units $j \neq i$, in Equation 1.

To illustrate this further, consider a field experiment investigating the effects of certain treatments on crop yields, as stylized in Figure A2. If we apply fertilizer to field 'C2', we may find local spillover effects to neighboring fields. This effect may occur directly ($x_i \to y_j$), e.g., when roots surpass field boundaries, or via a third mediator ($x_i \to z_j \to y_j$) when the treatment is dispersed and affects the (unmeasured) 'treatment' of neighbors, e.g. when fertilizer is washed away by rain. Yields themselves are behind global spillovers, which may occur, e.g., if wheat grows too large for its stalks and knocks others over, or if there is a form of communication to prompt or dissuade growth.[5] We will encounter spatial autocorrelation if there are third factors that feature a spatial pattern. This could be a hedge blocking sunlight and wind, the topography, or a plethora of other factors.

---

[5]Such mechanisms can be beneficial for the fitness of a plant. Trees that grow too densely undermine their overall fitness, while a crop that is flowering successfully can indicate favorable conditions to others.
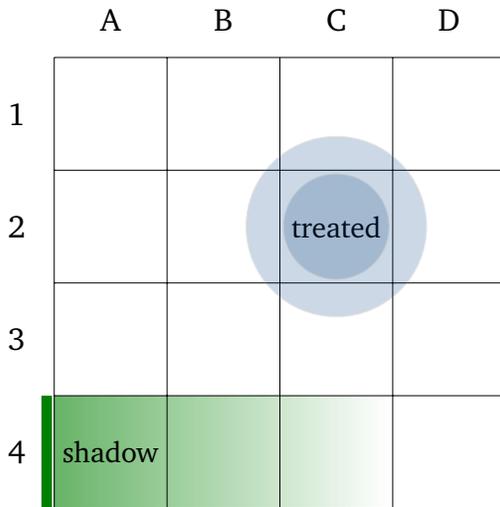
Figure A2: Stylised illustration of a grid of fields for analysing the effects of fertiliser on crop yield. The field in 'C2' is treated (e.g. with fertiliser), but spillover effects of the treatment (e.g. via dissolution) impact neighbouring fields. Meanwhile, a hedge to the left of field 'A4' causes correlation between it and its neighbours by throwing shade.
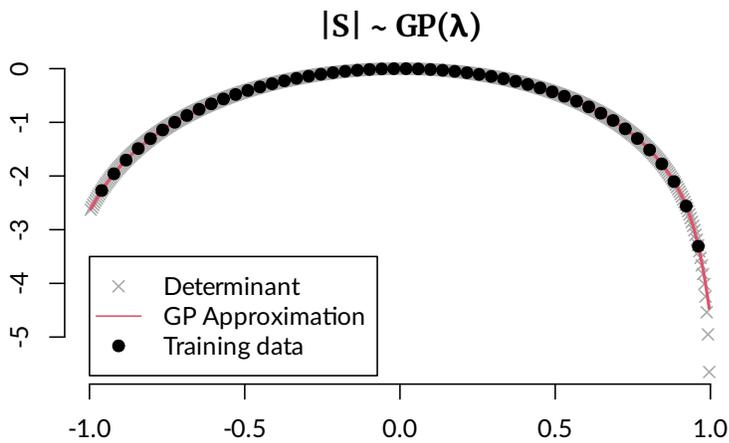
Figure A3: One-dimensional Gaussian process approximation to the Jacobian determinant $|\mathbf{S}(\lambda)|$ (on the vertical axis) using 50 training samples.
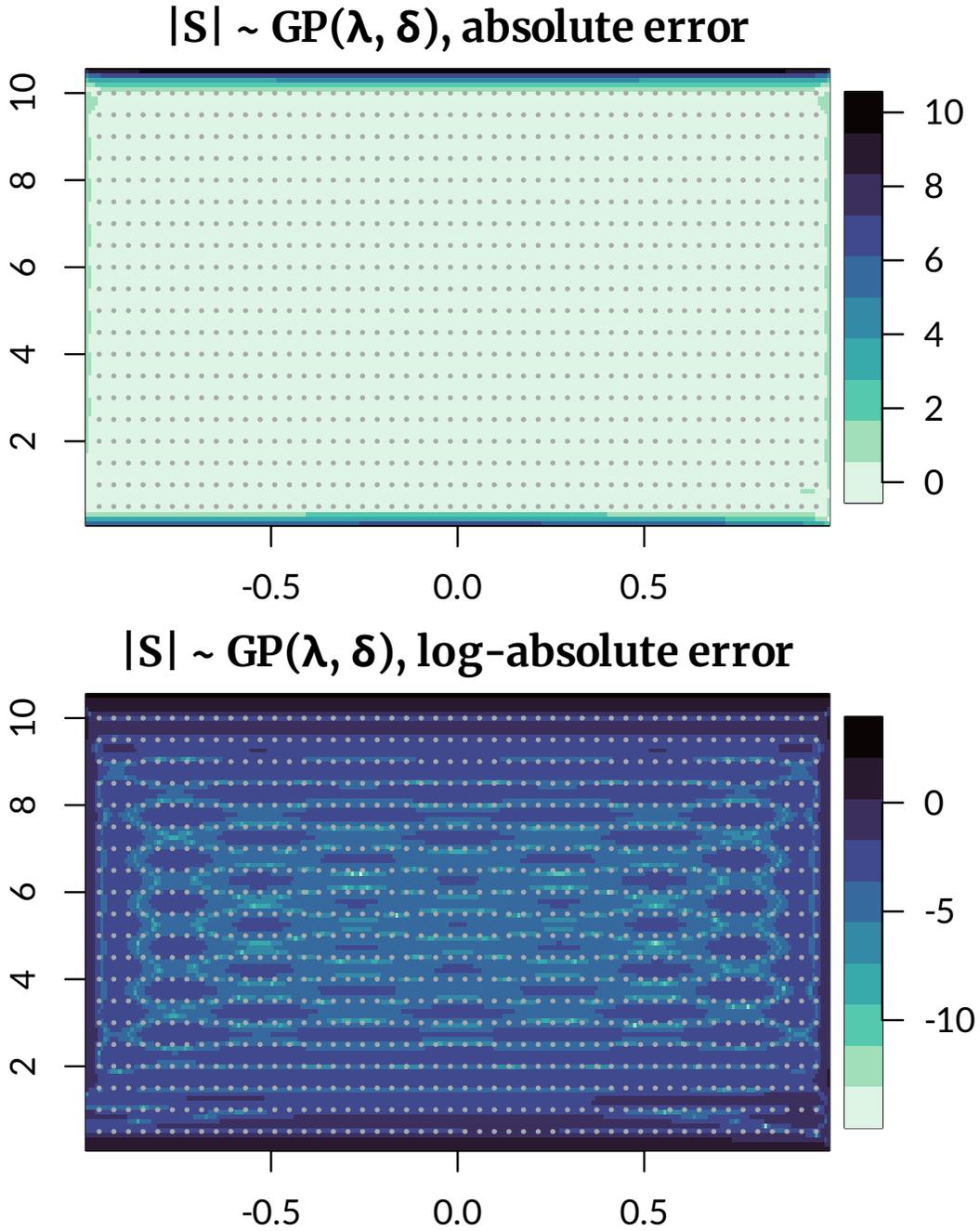
Figure A4: Absolute and log-absolute error of the Gaussian process approximation for the (two-dimensional) Jacobian determinant $|\mathbf{S}(\lambda, \delta)|$ using a 50×20 training grid. Distances are between $n = 100$ locations with coordinates sampled from a Uniform distribution. Gray dots indicate training samples.