Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatial Panel Models with Heterogeneous Coefficients: A Scalable, Integrated Hamiltonian Monte Carlo Estimation Framework

Yuheng LING

lyhneocc@gmail.com

Hainan Normal University

Peiyi Wu

Hainan Normal University

Lei JIANG

Guangzhou University

Research Article

Keywords: Bayesian spatial panel models, heterogeneous coefficients, hierarchical prior, Hamiltonian Monte Carlo, Hutchinson stochastic trace estimator

Posted Date: March 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8938359/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Geographical Systems on July 20th, 2026. See the published version at https://doi.org/10.1007/s10109-026-00508-4.

Spatial Panel Models with Heterogeneous Coefficients: A Scalable, Integrated Hamiltonian Monte Carlo Estimation Framework

Abstract

Spatial panel models with heterogeneous coefficients provide a flexible framework for capturing region-specific spillovers, but their practical use is often constrained by two challenges: (i) existing estimators primarily justified under large-T asymptotics, leaving unit-specific effects imprecisely estimated in the short panels common in regional science; and (ii) allowing spatial dependence and slopes to vary across units yields a high-dimensional parameter space that is difficult to estimate efficiently. This paper proposes a hierarchical Bayesian spatial model with heterogeneous coefficients that addresses both issues through hierarchical priors and an integrated Hamiltonian Monte Carlo estimation framework. The hierarchical prior induces partial pooling across units, shifting inference toward low-dimensional population hyperparameters while stabilizing unit-level estimates in finite samples, particularly when $N > T$. Furthermore, we develop scalable evaluation of the log-determinant term using a power-series expansion and Hutchinson stochastic trace estimation, and we exploit sparse matrix-vector multiplication and parallel computing to improve computational efficiency. Comprehensive simulation experiments demonstrate that the proposed Bayesian framework delivers reliable finite-sample inference and improved estimation accuracy relative to maximum likelihood in short panels. The results also show that heterogeneous spatial dependence is sensitive to the density of the spatial weights matrix than slope coefficients, offering practical guidance for applied research.

Keywords

Bayesian spatial panel models; heterogeneous coefficients; hierarchical prior; Hamiltonian Monte Carlo; Hutchinson stochastic trace estimator

JEL Classification: C11 · C13 · C23

1. Introduction

Spatial dependence has long been recognized as a fundamental feature of regional and spatial economic data. Economic outcomes observed for regions, cities, or countries are rarely independent across space. Instead, they are shaped by geographic proximity, economic integration and institutional linkages. Such spatial interdependencies give rise to spatial spillovers that propagate shocks beyond their point of origin, making spatial econometric models an essential tool in regional science. Over the past decades, spatial

autoregressive (SAR) models and their extensions have become popular specifications for analyzing these dependence structures in cross-sectional and panel data settings.

Conventional SAR panel specifications often rely on the assumption of a homogeneous spatial autoregressive parameter. In other words, the strength of spatial dependence is governed by a single scalar parameter that applies uniformly to all regions. While this assumption simplifies estimation and interpretation, it is often difficult to justify in practice. Regions differ markedly in their degree of openness, institutional integration, economic structure. As a result, the intensity with which economic shocks transmit across space is unlikely to be constant. Imposing homogeneous spatial interaction may therefore obscure important heterogeneity in spillover mechanisms and lead to misleading inference about the nature and magnitude of spatial effects.

Motivated by these concerns, a growing literature has focused on heterogeneous-coefficients spatial autoregressive (HSAR) panel data models that allow spatial autoregressive parameters as well as slope coefficients to vary across regions (Aquaro et al., 2021; Autant-Bernard & LeSage, 2019; LeSage & Chih, 2018; LeSage & Chih, 2016) These HSAR models represent an important conceptual advance, as they permit richer forms of spatial dependence and heterogeneous spillover effects, allowing researchers to move beyond average spatial spillovers. By accommodating region-specific spillover intensity, HSAR models provide a more realistic description of spatial processes in settings where integration, policy regimes, or economic linkages differ systematically across space.

Despite their conceptual appeal, existing HSAR models face several practical challenges. Most HSAR specifications are estimated using quasi-maximum likelihood (QML) methods and treat unit-specific spatial autoregressive parameters as fixed but unknown quantities. Theoretical justification for these

estimators typically relies on large-T asymptotic assumptions, which ensure identification and asymptotic normality of the heterogeneous parameters. However, such conditions are rarely satisfied in regional science applications, where the number of time periods T is limited as compared with the number of spatial units N in most spatial panels. As a result, inference in HSAR models may be unreliable when the time dimension is limited.

These inferential challenges are closely linked to computational and numerical difficulties arising from proliferation of the number of parameters. Allowing spatial autoregressive parameters to vary across regions substantially increases the dimensionality of the parameter space and complicates the optimization of the QML objective function. Moreover, treating unit-specific parameters as fixed effects limits the ability to borrow information across regions, thereby diminishing estimation efficiency and potentially weakening finite-sample performance. In a word, these issues have constrained the practical use of HSAR models in applied research, despite their advantages over homogeneous counterparts.

This study addresses these limitations by proposing an extended Bayesian HSAR model with hierarchical priors (hereafter referred to as the hierarchical Bayesian HSAR model (HHSAR)), estimated using an integrated Bayesian estimation framework based on Hamiltonian Monte Carlo (HMC). Our approach builds on two key ideas. First, we introduce hierarchical priors on unit-specific parameters to achieve partial pooling across units. Instead of estimating unit-specific parameters independently, we place hierarchical priors that induce partial pooling across units. This structure allows sharing information across units to stabilize estimation while still preserving unit-specific heterogeneity. As a result, the model relaxes the usual reliance on large-T asymptotics and improves inference in panels with small to moderate T. Second, we employ an integrated Bayesian estimation approach based on HMC. HSAR

models are characterized by high-dimensional and correlated posterior distributions, particularly when both slope coefficients and spatial dependence parameters vary across regions. Standard Gibbs or Metropolis-Hastings (H-M) algorithms often struggle in such context, requiring substantial tuning and long chains to achieve convergence. HMC, by contrast, exploits gradient information to efficiently explore complex posterior geometries and has proven effective in high-dimensional Bayesian models. By integrating HMC into the estimation framework of HHSAR models, we provide a practical and scalable approach that remains feasible even though the number of heterogeneous parameters is large.

To further enhance computational efficiency, we combine modern Bayesian sampling with state-of-the-art numerical techniques. In particular, we rely on power-series expansions and Hutchinson trace estimators to approximate the Jacobian (log-determinant) term arising from log-likelihood function. These approximations substantially reduce computational cost while retaining accuracy in the parameter regions relevant for posterior inference. In addition, we exploit sparse matrix-vector (SpMV) multiplication and parallel computing. All these techniques allow the proposed HHSAR model to be estimated within reasonable runtimes, making it feasible for applied research.

The contribution of this paper is threefold. First, we extend the existing HSAR model in the literature by introducing hierarchical priors that relax large-T. This extension is particularly relevant for regional science applications, where data constraints often preclude the use of asymptotic approximations underlying existing estimators. Second, we introduce HMC as a primary estimation tool for HHSAR models and show that the computational burden can be substantially alleviated by combining HMC with SpMV and parallel computing. Third, the simulation results provide clear evidence that, relative to maximum likelihood estimation, the proposed Bayesian approach delivers higher estimation accuracy in

small samples, benefiting from partial pooling and hierarchical priors.

Our analysis also sheds light that the proposed HMC-based approach can deliver reliable inference for model parameters through comprehensive simulation experiments. Meanwhile, it reveals that allowing spatial dependence to vary across units introduces additional sensitivity to different spatial weights matrices. In particular, the precision of spatial autoregressive parameters declines as spatial connectivity becomes denser, whereas slope estimates remain broadly stable across alternative weighting schemes. These findings clarify the interaction between heterogeneous spatial dependence and weights matrix construction, highlighting practical limitations for empirical applications.

The remainder of the paper is organized as follows. Section 2 introduces the HSAR panel model and outlines the proposed HHSAR specification. Section 3 describes the HMC-based integrated Bayesian approach and the computational techniques used to efficiently evaluate the model. Section 4 presents simulation results that assess the finite-sample performance of the proposed approach under a variety of panel dimensions and spatial structures. Section 5 provides an empirical illustration, and Section 6 concludes with a discussion of implications and directions for future research.

2. A heterogeneous-coefficient spatial autoregressive panel model

Aquaro et al. (2015) first propose a so-called HSAR panel model. Unlike conventional SAR panel models that impose a scalar spatial autoregressive parameter, this model allows the strength of spatial dependence to vary across cross-sectional units as well as introduces unit-specific slope parameters, yielding a flexible, full heterogeneous-coefficient SAR panel specification. Estimation of such model can be carried out using ML estimation and GMM pertaining to the frequentist approach (Aquaro et al., 2021). In addition, LeSage and Chih (2018) (LC hereafter) then propose the Bayesian version of this model and employ the H-M algorithm to estimate model parameters.

2.1 Model specification

Let y_{it} denotes the dependent variable observed for spatial unit $i = 1, \dots, N$ at time period $t = 1, \dots, T$, and let x_{it}^k be exogenous regressors. The heterogeneous-coefficients SAR panel model can be written as

$$y_{it} = \alpha_i + \psi_i \sum_{j=1}^{N} w_{ij}\, y_{jt} + \sum_{k=1}^{K} \beta_i^k x_{it}^k + \varepsilon_{it},$$

where w_{ij} denotes the (i,j)-th element of a known spatial weights matrix W. ψ_i is the unit-specific spatial autoregressive parameter, and β_i^k is the k-th $N \times 1$ vector of unit-specific regression coefficients.

Stacking observations across units for each time period t yields the matrix form

$$\mathbf{y}_t = \boldsymbol{\Psi} W \mathbf{y}_t + \mathbf{B}\mathbf{z}_t + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{y}_t = (y_{1t}, \dots, y_{Nt})'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$. $\boldsymbol{\Psi} = \mathrm{diag}(\psi_1, \dots, \psi_N)$ is a diagonal matrix collecting unit-specific spatial autoregressive parameters. To accommodate heterogeneous intercepts and slope coefficients, the stacked regressor vector is define as $\mathbf{z}_t = (\boldsymbol{\iota}', \mathbf{x}_t^{(1)\prime}, \dots, \mathbf{x}_t^{(K)\prime})'$, where $\boldsymbol{\iota} = \mathbf{1}_N$ and $\mathbf{x}_t^{(k)} = (x_{1t}^k, \dots, x_{Nt}^k)'$. The corresponding coefficient matrix is given by

$$\mathbf{B} = [\boldsymbol{\alpha}, B^1, \dots, B^K],$$

with $\boldsymbol{\alpha} = \mathrm{diag}(\alpha_1, \dots, \alpha_N)$ and $B^k = \mathrm{diag}(\beta_1^k, \dots, \beta_N^k)$.

Provided that the matrix $I_N - \boldsymbol{\Psi} W$ is nonsingular, the reduced form is given by

$$\mathbf{y}_t = (I_N - \boldsymbol{\Psi} W)^{-1}(\mathbf{B}\mathbf{z}_t + \boldsymbol{\varepsilon}_t).$$

As in LC, this condition is ensured by constraining $\psi_i \in (\frac{1}{\lambda_{min}}, \frac{1}{\lambda_{max}})$ under the W transformed to its doubly stochastic form. In contrast to homogeneous SAR models, spatial spillovers in this setting propagate heterogeneously, allowing for asymmetric spill-in and spill-out dynamics that differ across units (LeSage & Chih, 2016).

The corresponding log-likelihood function of the HSAR model takes the following form,

$$\ell(\theta) = T \log | I_N - \boldsymbol{\Psi} W | - \frac{1}{2} \sum_{t=1}^{T} (\mathbf{y}_t - \boldsymbol{\Psi} W \mathbf{y}_t - \mathbf{B}\mathbf{z}_t)^{\top} \Omega^{-1} (\mathbf{y}_t - \boldsymbol{\Psi} W \mathbf{y}_t - \mathbf{B}\mathbf{z}_t),$$

where $\Omega = diag(\sigma_1^2, \ldots \sigma_N^2)$ and the parameter vector to be estimated is $(\beta_1^1 \ldots \beta_N^1, \ldots, \beta_1^k \ldots \beta_N^k, \psi_1, \ldots, \psi_N, \sigma_1^2, \ldots \sigma_N^2)$.

The HSAR model introduced above entails a high-dimensional parameter space and a non-standard likelihood, rendering computational challenges for Bayesian estimation, particularly in panels with large cross-sectional dimensions. Bayesian inference requires repeated evaluation of the log-likelihood, involving the Jacobian term $\log | I_N - \boldsymbol{\Psi} W |$ as well as unit-specific regression coefficients β_i^k, whose dimensionality increases linearly with the number of spatial units. To address these challenges, we develop a HHSAR specification by extending the HSAR model to incorporate hierarchical priors for heterogeneous coefficients. This hierarchical structure induces partial pooling across units and regularizes the high-dimensional parameter space (Gelman, 2006).

2.2 Bayesian prior information

In LC's Bayesian procedure, regression coefficients β_i^k and spatial autoregressive parameters ψ_i are assigned independent Gaussian priors centered on common prior means, while independent region-specific disturbance variances follow inverse-gamma distributions. The prior means are often interpreted as empirical Bayes inputs obtained from homogeneous spatial panel estimates, allowing the degree of parameter heterogeneity to be controlled through the choice of prior variances. The theoretical admissibility of the spatial autoregressive parameters is not enforced through the support of prior distributions. Instead, the stability condition $-1 < \psi_i < 1$ is imposed during MCMC estimation by rejecting conditional posterior draws that fall outside the admissible region. This strategy is natural in Gibbs sampling or M-H algorithms, where parameters are updated sequentially and parameter constraints can be enforced locally at each step.

Beyond computational convenience, this prior construction also has substantive implications for spatial panels. In LC, heterogeneity is operationalized by assigning unit-specific parameter vectors $\mathbf{B}$ and placing independent Gaussian priors on $\{\beta_i^k\}_{i=1}^N$, treating these coefficients as unrelated fixed but unknown quantities. Under this formulation, posteriors of $\{\beta_i^k\}_{i=1}^N$ are driven primarily by the within-unit time series of length T. Holding T fixed, adding more cross-sectional units provides little direct information about any particular β_i^k. Consequently, under the common spatial panel asymptotic, $N \to \infty$ with fixed T, the model effectively includes N unit-specific parameters, each supported by only T observations, which could induce finite-T bias, akin to the incidental-parameter problem.

Our specification, however, adopts a hierarchical prior (Chib, 2008) that delivers two practical advantages for spatial panels (often $N > T$). First, it induces partial pooling: unit-level coefficients are shrunk toward a common population distribution, so information is shared across units while still allowing heterogeneity. This Bayesian shrinkage regularizes weakly identified unit effects, reducing variance inflation and improving finite-sample stability when T is short. Second, it shifts identification from N unrelated coefficients to a low-dimensional set of hyperparameters. As N grows, the cross section primarily identifies the population moments, while unit-specific effects are learned conditionally given these moments through partial pooling (McElreath, 2018). Therefore, even under fixed T and $N \to \infty$, the hyperparameters become increasingly well-informed, and the hierarchical prior mitigates the finite-T bias.

Additionally, this hierarchical prior is deliberately chosen to accommodate HMC. HMC performs well under a jointly parameterized model with posterior geometry that is well-behaved in high dimension. To this end, we avoid independent, unit-by-unit prior specifications and instead model the heterogeneous

coefficients as conditionally exchangeable draws from a common multivariate distribution[1]. For each of the N units, its $K \times 1$ vector of coefficients is modeled as

$$\beta_i \mid \bar{\beta}, \boldsymbol{\Sigma_\beta} \sim \mathcal{N}\left(\bar{\beta}, \boldsymbol{\Sigma_\beta}\right), i = 1, \ldots, N,$$

where $\bar{\beta}$ (a $K \times 1$ vector) is the global mean and $\boldsymbol{\Sigma_\beta}$ (a $K \times K$ matrix) is the global covariance. These hyperparameters are endowed with priors and learned jointly from the full panel. This joint Bayesian parameterization yields a smoother posterior surface and more coherent global curvature than sequentially constrained updates, improving numerical stability and sampling efficiency in HMC estimation.

Moreover, to facilitate efficient HMC sampling, we adopt the non-centered parameterization of the hierarchical prior. Specifically, $\boldsymbol{\Sigma_\beta}$ is decomposed as

$$\boldsymbol{\Sigma}_\beta = \mathrm{diag}(\boldsymbol{\tau})\, \mathbf{R}_{\boldsymbol{\beta}}\, \mathrm{diag}(\boldsymbol{\tau}),$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K)$ are marginal standard deviations and $\mathbf{R}_{\boldsymbol{\beta}}$ is a correlation matrix. The unit-specific coefficients are then expressed as

$$\beta_i = \bar{\beta} + \mathrm{diag}(\boldsymbol{\tau})\mathbf{L}_{\boldsymbol{\beta}}\mathrm{Z}_i, \mathrm{Z}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_K),$$

with $\mathbf{L}_{\boldsymbol{\beta}}$ denoting the Cholesky factor[2] of $\mathbf{R}_{\boldsymbol{\beta}}$.

Weakly informative Gaussian priors are assigned to the global mean coefficients,

$$\bar{\beta} \sim \mathcal{N}(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}_K),$$

while half-normal priors are placed on the marginal parameter τ_k. For the correlation structure $\mathbf{R}_{\boldsymbol{\beta}}$, we adopt an LKJ Cholesky covariance prior[3],

---

[1] First, it induces partial pooling across spatial units: estimates for units with limited information (e.g., short time span) are shrunk toward the global mean $\bar{\beta}$, which stabilizes inference and mitigates finite-T biases. Second, the model efficiently shares information across units, because data from all spatial units contribute to estimating the hyperparameters $\bar{\beta}$ and $\boldsymbol{\Sigma_\beta}$, which in turn inform each unit-specific $\{\beta_i^k\}_{i=1}^N$. Third, the degree of heterogeneity is adaptively learned from the data through $\boldsymbol{\Sigma_\beta}$. A small estimated covariance implies strong homogeneity, while a larger covariance allows substantial regional flexibility. Finally, this hierarchical prior is more parsimonious than specifying separate priors for each unit, as it requires only a single mean vector and covariance matrix to describe the entire distribution of regional coefficients.

[2] $\mathbf{R}_{\boldsymbol{\beta}} = \mathbf{L}_{\boldsymbol{\beta}}\mathbf{L}_{\boldsymbol{\beta}}^T$

[3] Proposed by Lewandowski et al. (2009), this Lewandowski-Kurowicka-Joe (LKJ) covariance prior is particularly well suited for

$$\mathbf{R}_{\boldsymbol{\beta}} \sim \text{LKJ}(\zeta),$$

with $\zeta > 1$, which mildly favors lower correlations while remaining weakly informative.

For the disturbance, we place a half-Student-t prior on the standard deviation,

$$\sigma \sim t^{+}(0,1),$$

rather than an inverse-gamma prior on σ^2. While inverse-gamma priors are convenient in conditionally conjugate Gibbs samplers, they are known to induce unfavorable posterior geometry in gradient-based samplers, particularly in models where the variance parameter interacts strongly with nonlinear components of the likelihood. The half-Student-t prior provides a weakly informative and computationally stable alternative that is now standard in modern Bayesian practice (Gelman et al., 2017).

The heterogeneous spatial autoregressive parameters are parameterized as

$$\psi_i = \tanh(\eta_i), \eta_i|\sigma_\eta \sim \mathcal{N}\left(0, \sigma_\eta^2\right),$$

where $\tanh(\cdot)$ denotes the hyperbolic tangent transformation. We sample the unconstrained latent variable $\eta_i \in \mathbb{R}$ and enforce stability through this reparameterization, which guarantees $\psi_i \in (-1,1)$ throughout the HMC trajectory and removes the need for explicit boundary handling. Sampling on the unconstrained η_i-scale also improves efficiency relative to sampling ψ_i, directly, since posterior geometry typically becomes highly curved near ± 1, impeding exploration.

Moreover, the elicited prior on ψ_i is symmetric around zero and assigns diminishing probability mass near ± 1, providing mild regularization against near-instability. This feature is particularly important in the present setting, since extreme ψ_i values could generate sharp curvature in the likelihood and exacerbate numerical error in stochastic log-determinant approximations. Importantly, the substantive interpretation of heterogeneous spatial autoregressive parameters is unchanged relative to LC; we simply

HMC estimation, as it yields smooth posterior geometry and avoids pathological behavior near the boundaries of the correlation space.

adopt a parameterization that is better suited to joint, gradient-based Bayesian estimation. In the next section, we will elaborate on an HMC-based integrated Bayesian approach tailored to our HHSAR model.

3. Bayesian estimation strategy

The proposed Bayesian estimation framework can handle both the number of spatial units and the degree of parameter heterogeneity are large, while maintaining the scalability. The nutshell of the estimation framework can be divided into two core aspects. First, posterior inference is conducted using HMC (Neal, 2011), which is well suited to hierarchical parameterizations. Second, to overcome the principal computational bottleneck associated with evaluating the log-determinant $\log|I_N - \boldsymbol{\Psi} W|$, we adopt a stochastic approximation based on the Hutchinson trace estimator combined with a truncated power-series expansion (Han et al., 2015).

3.1 HMC method

HMC represents a fundamental advance in statistical computation, replacing the random-walk behavior of traditional Markov Chain Monte Carlo methods with physics-informed, gradient-guided exploration of probability distributions (Betancourt, 2019). While conventional M-H algorithm performs random walks through parameter space, which can become cripplingly slow in high dimensions. HMC leverages the geometry of posterior distributions to make coherent, long-distance proposals. This capability makes it particularly attractive for complex models.

The Hamiltonian dynamics underlying HMC transforms the posterior sampling problem into the physical simulation of a frictionless puck moving across an energy landscape (Betancourt et al., 2017). The total Hamiltonian $H(\theta, r) = U(\theta) + K(r)$ combines potential energy $U(\theta) = -log\pi(\theta|data)$ with kinetic energy $K(r)$ from auxiliary momentum variables r. Critically, the marginal distribution of parameters θ under $exp(-H(\theta, r))$ is exactly the target posterior. The system evolves according to Hamilton's equations, which preserve both total energy and phase space volume, enabling efficient

navigation of the typical set, i.e., the region of high probability mass in high-dimensional spaces. The leapfrog integrator discretizes these dynamics while maintaining the symplectic (volume-preservation and time-reversibility) property essential for high acceptance probabilities (Betancourt et al., 2017).

An HMC iteration follows three steps (Mukherjee & Vats, 2026).

1. Momentum Resampling. A new momentum vector r is drawn from a Gaussian distribution that is independent of θ. This randomizes the direction of exploration.

2. Hamiltonian Simulation. Starting from the current state (θ, r), the Hamiltonian dynamics are simulated for a fixed trajectory length using a leapfrog integrator. This discretized, symplectic scheme approximates the continuous dynamics through alternating half-steps in momentum and full steps in position. Crucially, it requires only the gradient $\nabla U(\theta)$, not the posterior's normalizing constant, enabling efficient computation via automatic differentiation. By following the terrain's gradients, it proposes a new state (θ', r') far from the starting point while maintaining a high probability of acceptance.

3. Metropolis Acceptance. The proposed state $\theta'$is accepted with probability $min(1, exp(-H(\theta', r') + H(\theta, r)))$. This corrects for numerical error from the discretized simulation, guaranteeing exact sampling from the target.

For our HHSAR model, this physics-inspired HMC method offers decisive computational advantages. The posterior density, proportional to the product of the likelihood and specified priors, presents particular challenges through the Jacobian term $\log |I_N - \boldsymbol{\Psi} W|$, which enters additively with multiplicity T. Unlike MCMC approaches that rely on conditional conjugacy for subsets of parameters

and M-H updates for spatial autoregressive parameters strategies that may experience from slow mixing and require extensive fine-tuning in panels, HMC jointly updates all parameters using gradient information.

This includes the entire N-dimensional vector of heterogeneous spatial parameters within a single HMC transition, eliminating the need for blockwise sampling or carefully tuned proposal distributions. The algorithm's efficiency stems from its ability to follow the posterior geometry along coherent trajectories determined by Hamilton's equations, dramatically improving mixing properties in high-dimensional parameter spaces. Stan's implementation through the No-U-Turn Sampler (Hoffman & Gelman, 2014) automates the critical tuning of step size and trajectory length, making this powerful approach accessible while maintaining theoretical guarantees.

Furthermore, the practical implementation of HMC for HHSAR models also leverages several innovations. First, automatic differentiation efficiently computes gradients of the complete posterior. Second, parallelization through Stan's built-in function ($reduce_sum$) (Carpenter et al., 2017) distributes computation of stochastic trace estimators across probe vectors when approximating the Jacobian term, allowing computational cost to grow sub-linearly with the number of probes. The combination of gradient-based exploration of HMC, Hutchinson trace estimation, and computational parallelism makes full Bayesian inference feasible for large N panels, enabling estimation that scales effectively with both many spatial units and long time periods.

3.2 Log-Determinant Approximation via Hutchinson-series approximation

In spatial econometric literature, a common computational bottleneck arises from evaluating the Jacobian term in the log-likelihood function, as spatial dependence in the dependent variable results in a non-identity Jacobian matrix. This issue arises directly in the HHSAR model with the following

Jacobian term:

$$\sum_{t=1}^{T} \log \mid I_N - \boldsymbol{\Psi} W \mid = T \cdot \log \mid I_N - \boldsymbol{\Psi} W \mid.$$

The Gaussian likelihood involves the log-determinant of a transformation matrix of the form $I_N - \boldsymbol{\Psi} W$, where W is usually a sparse spatial weights matrix. Even though W is sparse, the matrix $I_N - \boldsymbol{\Psi} W$ could be dense in general, and its determinant cannot be computed efficiently by standard linear algebra routines when N is large. We therefore use a scalable approximation based on a power-series expansion coupled with a stochastic trace estimation.

3.2.1 Power-series expansion of $\log \mid I_N - \boldsymbol{\Psi} W \mid$

A key is that log-determinants can be expressed in terms of matrix traces. Under the condition that the spectral radius satisfies $\rho(\boldsymbol{\Psi} W) < 1$, we have the matrix identity by letting $\mathrm{A} = \boldsymbol{\Psi} W$,

$$\log \mid I_N - \mathrm{A} \mid = \mathrm{tr}[\log (I_N - \mathrm{A})] = -\Sigma_{k=1}^{\infty} \frac{\mathrm{tr}[\mathrm{A}^k]}{k}.$$

Thus, approximating $\log \mid I_N - \mathrm{A} \mid$ reduces to computing traces of powers of A. In practice, we truncate at order K_{trunc}:

$$\log \mid I_N - \mathrm{A} \mid \approx \Sigma_{k=1}^{K_{\text{trunc}}} \frac{\mathrm{tr}[\mathrm{A}^k]}{k}.$$

The approximation error decreases with larger K_{trunc}, and it decreases more rapidly when $\rho(\mathrm{A})$ is well below 1. With row-standardized W, constraining $\psi_i \in (-1,1)$ and avoiding values near ± 1 typically helps maintain $\rho(\mathrm{A}) < 1$, improving both convergence and numerical stability.

3.2.2 Hutchinson trace estimator

The remaining challenge is evaluating $\mathrm{tr}[\mathrm{A}^k]$ for $k = 1, \dots, K_{\text{trunc}}$. However, direct evaluation of $\mathrm{tr}[\mathrm{A}^{K_{\text{trunc}}}]$ would require dense matrix operations. Hutchinson (1989) introduced an elegant, randomized estimator that bypasses explicit matrix construction and relies only on matrix-vector products. Let $z \in \{-1, +1\}^N$ be a Rademacher random vector (Avron & Toledo, 2011) with independent entries. Then

$$\mathbb{E}[z^\top B z] = \mathrm{tr}(B)$$

for any square matrix B. Using R independent probe vectors z_r, We approximate

$$\mathrm{tr}(B) \approx \frac{1}{R}\sum_{r=1}^{R} z_r^{\top} B z_r.$$

Each term $z_r^{\top} B z_r$ requires only a single multiplication of B with a vector, making the trace estimator computationally feasible for large matrix B is only available implicitly. More importantly, since each trace $\mathrm{tr}[\mathrm{A}^k]$ is approximated via the Hutchinson estimator, the powers of $\boldsymbol{\Psi W}$ are never formed explicitly. In addition, for each probe vector z_r, we compute a recurrence

$$v_{r,0} = z_r, v_{r,k} = \boldsymbol{\Psi} W v_{r,k-1},$$

so that $v_{r,k} = (\boldsymbol{\Psi} W)^k z_r$ and $z_r^{\top} v_{r,k} = z_r^{\top} (\boldsymbol{\Psi} W)^k z_r$. Each iteration requires only a SpMV multiplication with W, followed by elementwise scaling by $\boldsymbol{\Psi}$.

Finally, the resulting approximation gives

$$log\ \widehat{det\,(I - \boldsymbol{\Psi} W)} \approx -\Sigma_{k=1}^{K_{\mathrm{trunc}}} \frac{1}{k}\left[\frac{1}{R}\sum_{r=1}^{R} z_r^{\top} (\boldsymbol{\Psi} W)^k z_r\right].$$

The overall accuracy of the approximation is controlled by the two parameters, the number of probe vectors R and the truncation order K_{trunc}. Approximation error tends to worsen as $\max|\psi_i|$ approaches 1, motivating mild prior regularization away from the boundary.

3.2.3 Computational properties and implementation

The computational cost of the approximation scales as

$$O(R \cdot K_{\mathrm{trunc}} \cdot \mathrm{nnz}(W)),$$

where $\mathrm{nnz}(W)$ is the number of nonzero elements in the spatial weights matrix. This makes the approach well suited to large spatial systems. In the Stan implementation, the computation involving multiple probe vectors is distributed across available processor cores through parallel processing (Davis & Vaddady, 2025), while SpMV multiplication is carried out using compressed sparse row format routines (Gao et al., 2024). Together, these designs enable efficient Bayesian inference for the HHSAR models incorporating heterogeneous spatial autoregressive parameters and unit-specific regression coefficients.

This integrated approach offers several advantages over conventional MCMC algorithms, including (1) superior mixing properties mitigating random-walk inefficiency, (2) joint updating avoiding conditional dependencies that slow block-wise sampling, (3) automated tuning procedures reducing researcher intervention, and (4) scalable computation enabling by parallelization of expensive matrix operations. By transforming how we explore the complex posterior distributions arising in spatial econometrics, HMC has expanded the scope of models amenable to full Bayesian inference while providing practical computational efficiency for large-scale applications.

4. Monte Carlo simulation

In this section, we evaluate the performance of the HMC-based integrated Bayesian approach for estimating HHSAR panel models using Monte Carlo (MC) simulation studies. Specifically, we consider the following experimental designs:

(i) given the rock-contiguity weights matrix W, we vary the sample size by considering different combinations of N and T;

(ii) for a fixed sample size, we examine alternative specifications of spatial weights matrices;

(iii) we compare the HMC approach with the classical frequentist ML estimator.

We end this MC study by reporting the computational efficiency of the HMC approach under different experimental settings. The goal of these simulations is to validate the proposed estimation approach, demonstrating that it delivers reliable parameter estimates and valid statistical inference, and that, under weakly informative priors, it produces results comparable to QML estimation while maintaining favorable computational performance.

We consider the following data-generating process (DGP):

$$y_{it} = \alpha_i + \psi_i \sum_{j=1}^{N} w_{ij}\, y_{jt} + \beta_i^{(1)} x_{it}^{(1)} + \beta_i^{(2)} x_{it}^{(2)} + \varepsilon_{it},$$

which includes two exogenous regressors and a heterogenous intercept α_i. The vector of unknown parameters is

$$\boldsymbol{\theta} = \{\boldsymbol{\Psi}, \mathbf{B}, (\sigma_1^2, \dots \sigma_N^2)^T\},$$

resulting in a total of $5N$ unknown parameters.

The regressors $x_{it}^{(1)}$and $x_{it}^{(2)}$ are generated independently from $N(0,1)$ and $N(1,1)$, respectively. The heterogeneous intercepts follow $\alpha_i \sim N(\alpha_\mu, \sigma_{\tau\alpha}^2)$ with $\alpha_\mu = 0.2$, $\alpha_{\tau\alpha} = 0.3$. Regression coefficients are generated as $\beta_{i1} \sim N(\beta_{\mu1}, \sigma_{\tau1}^2)$ with $\beta_{\mu1} = 0.6$, $\sigma_{\tau1} = 0.25$, and $\beta_{i2} \sim N(\beta_{\mu2}, \sigma_{\tau2}^2)$ with $\beta_{\mu2} = 0.3$, $\sigma_{\tau2} = 0.15$. Spatial autoregressive parameters are specified as $\psi_i = \tanh(\eta_i)$, $\eta_i \sim N(0, \psi_{\text{range}}^2)$ with $\psi_{\text{range}} = 0.5$, ensuring admissibility of spatial autoregressive coefficients. All hyperparameters are fixed across experiments. The idiosyncratic errors are generated as Gaussian disturbances with heterogeneous variances, $\sigma_i^2 \sim N(0,1)$.

For the spatial weights matrix $W = (w_{ij})$, we first construct a non-normalized matrix using rock contiguity criteria and then transform it into a doubly stochastic matrix via the Sinkhorn-Knopp algorithm (Knight, 2008). In addition, we consider alternative neighbor definitions, including queen contiguity criteria, k-nearest-neighbors with varying numbers of neighbors and distance-based truncation schemes. These specifications yield spatial weight matrices with differing degrees of sparsity. Given the DGP above, the outcome variable is generated as[4] $y_t = (I_N - \boldsymbol{\Psi} W)^{-1}(\mathbf{B}\mathbf{z}_t + \boldsymbol{\varepsilon}_t)$.

[4] Stacking observations by units yields

$$\mathbf{y}_t = \boldsymbol{\Psi} W \mathbf{y}_t + \mathbf{B}\mathbf{z}_t + \boldsymbol{\varepsilon}_t.$$

where $\boldsymbol{\Psi} = \text{diag}(\psi_1, \dots, \psi_N)$, $\mathbf{B}$ collects the heterogeneous regression coefficients, and $\mathbf{z}_t$ stacks all regressors.

To assess performance under both small and large cross-sectional dimensions, we consider three values of $N \in \{25, 50, 100\}$ and four values of $T \in \{15, 25, 50, 100, 125\}$, covering cases where $N < T$, $N = T$, and $N > T$. For each (N, T)configuration, we conduct $M = 500$ Monte Carlo replications.

Simulation results are summarized using three performance measures, including averaged empirical bias (Bias), root mean squared error (RMSE) (Aquaro et al., 2021), empirical standard deviation (E-SD), and empirical coverage probabilities (Coverage) (Tan & Elhorst, 2023). Bias, averaged across units, is computed as

$$Bias = (NM)^{-1} \sum_{i=1}^{N} \sum_{r=1}^{M} \left(\hat{\gamma}_{i,r} - \gamma_{i,0}\right),$$

where $\hat{\gamma}_{i,r}$ denotes the estimated parameter $\gamma \in \{\psi_i, \beta_i^1, \beta_i^2, \sigma_i\}$ in replication r, and $\gamma_{i,0}$ is the true value.

The RMSE, averaged across units, is calculated as

$$RMSE = N^{-1} \sum_{i=1}^{N} \sqrt{M^{-1} \sum_{r=1}^{M} \left(\hat{\gamma}_{i,r} - \gamma_{i,0}\right)^2},$$

which reflects both systematic bias and sampling variability.

The empirical standard deviation (E-SD) measures the dispersion of the estimator across Monte Carlo replications and is defined as

$$E\text{-}SD = N^{-1} \sum_{i=1}^{N} \sqrt{\frac{1}{M-1} \sum_{r=1}^{M} \left(\hat{\gamma}_{i,r} - \bar{\gamma}_i\right)^2},$$

where $\bar{\gamma}_i = M^{-1} \sum_{r=1}^{M} \hat{\gamma}_{i,r}$.

In addition to these metrics, we assess the accuracy of uncertainty quantification using empirical coverage probabilities. For each replication, a 95% credible interval (CI) is constructed for every parameter based on the retained 800 post-warmup HMC draws[5]. Coverage is defined as the proportion

[5] An initial burn-in of 400 iterations is discarded.

of simulations in which the true parameter value lies within this interval. Under correct model calibration, a nominal 95% interval should contain the true parameter in approximately 95% of the replications. With 500 replications, this corresponds to roughly 475 intervals covering the true value. Deviations from this benchmark indicate over- or under-estimation of posterior uncertainty and directly reflect the reliability of statistical inference.

4.1. Finite sample properties of integrated Bayesian estimation approach

Table 1 Monte Carlo results for the integrated Bayesian estimation approach

Metrics	Parameters	N	25			50			100		
		T	10	25	50	25	50	75	25	50	75
Bias	$\beta_i^{(1)}$		-0.0031	0.0012	0.0006	0.0001	-0.0003	0.0003	0.0002	-0.0004	0.0004
	$\beta_i^{(2)}$		-0.0033	0.0001	-0.0007	0.0019	0.0008	0.0002	0.0007	0.0003	-0.0005
	ψ_i		-0.0038	0.0042	0.0038	-0.0040	-0.0021	0.0015	0.0010	-0.0014	-0.0011
E-SD	$\beta_i^{(1)}$		0.1676	0.1176	0.0849	0.1159	0.0846	0.0692	0.0837	0.06	0.0539
	$\beta_i^{(2)}$		0.1248	0.0959	0.0757	0.0947	0.0748	0.0636	0.0741	0.0561	0.0511
	ψ_i		0.2777	0.1999	0.1486	0.2024	0.1511	0.1263	0.1573	0.1151	0.1042
Coverage	$\beta_i^{(1)}$		0.9471	0.9484	0.9503	0.9483	0.9479	0.9488	0.9501	0.9503	0.951
	$\beta_i^{(2)}$		0.9581	0.953	0.9601	0.9572	0.9564	0.9561	0.9573	0.9571	0.958
	ψ_i		0.9417	0.9481	0.9479	0.9503	0.9489	0.960	0.9488	0.951	0.9501
RMSE	$\beta_i^{(1)}$		0.1736	0.1182	0.0864	0.1177	0.0850	0.0696	0.0839	0.0599	0.0543
	$\beta_i^{(2)}$		0.1306	0.0999	0.0773	0.0985	0.0761	0.0642	0.0754	0.0564	0.0509
	ψ_i		0.3116	0.2794	0.2596	0.2450	0.2164	0.1678	0.1940	0.1666	0.1615

Table 1 reports Monte Carlo results for the proposed approach across different combinations of N and T. Across all panel configurations, our approach exhibits negligible bias for all heterogeneous parameters. The average biases remain close to zero in all scenarios. This indicates that the proposed approach successfully recovers the true parameter values under the specified DGP.

Precision improves markedly as the time dimension increases. For example, when $N = 25$, E-SD declines from 0.1676 at $T = 10$ to 0.0849 at $T = 50$, with the corresponding RMSE falling from 0.1737 to 0.0864. A similar monotonic decline is observed for larger N. Under $N = 100$, the largest RMSE occurs in the smallest T case (0.194), but decreases steadily to 0.1666 and 0.1615 as T increases. These reductions confirm that gains in accuracy stem primarily from improved precision rather than changes in bias.

Increasing the cross-sectional dimension also enhances finite sample performance. Holding T fixed, E-SD and RMSE decline moderately as N increases, although the gain is likely less pronounced than enlarging T. This behavior is consistent with the identification mechanism induced by the hierarchical prior. As N grows, the cross section can provide more robust unit-specific effects estimation through partial pooling.

Empirical coverage probabilities remain close to the nominal 95% level across all designs. For instance, coverage is 0.9417-0.9601 under $N = 25$ and remains within 0.9479-0.958 even in larger N cases. Importantly, no systematic under-coverage is observed, indicating well-calibrated posterior uncertainty.

In a word, Table 1 demonstrates that the HMC-based integrated Bayesian approach can deliver reliable finite-sample inference for the HHSAR specification. Even in short panels, bias remains negligible and coverage is well calibrated, while precision improves steadily with increases in T and, to a lesser extent,

N. The combination of hierarchical priors and HMC sampling can yield accurate estimates across a broad range of panel dimensions.

4.2. Impacts of weights matrix sparsity

Table 2 reports the Monte Carlo results of Experiment 2 under alternative spatial weight matrices under $N = 50$ and $T = 25$. Compared with the rook-contiguity W matrix used in Experiment 1, the current setting considers three commonly adopted spatial neighbor definitions in applied econometrics: a queen-contiguity (W_{queen}), a k-nearest neighbour (k-nn) with $k = 9$ (W_{knn}), as well as an inverse-distance based W matrix truncated at the 20 nearest neighbours (W_{Inv}). According to Hoyer's definition[6] (2004), the degree of sparsity decreases monotonically from W_{queen} ($\text{S}_{\text{parsity}}$ =0.65) to W_{knn} ($\text{S}_{\text{parsity}}$ =0.55) and further to W_{Inv} ($\text{S}_{\text{parsity}}$ =0.45), indicating progressively richer spatial connections relative to the rook one used previously ($\text{S}_{\text{parsity}}$ =0.75). The Sinkhorn-Knopp algorithm is applied to all W matrices to obtain the doubly stochastic form.

In contrast to the slope parameters, the estimation of ψ_i appears more sensitive to the density of W. As W becomes progressively denser, both bias and RMSE for ψ_i increase monotonically. Although the bias remains small in absolute magnitude, the accompanying increase in E-SD suggests some loss of estimation precision under denser spatial structures. Interestingly, the empirical coverage probability improves slightly as W becomes denser, suggesting increased posterior uncertainty rather than systematic bias.

When compared with results reported in LC, the magnitude of bias obtained under contiguity-based W in our approach is smaller in the scenarios considered. For example, when $N = 50$ and $T = 25$, LC

[6] $\text{S}_{\text{parsity}} = \frac{\sqrt{n} - \sum_i |x_i| / \sqrt{\sum_i x_i^2}}{\sqrt{n} - 1}$

reports an average bias of 0.0683 for ψ_i, whereas in our experiments the corresponding bias remains close to zero, with values of 0.0005 and 0.0008 under two alternative contiguity-based W. Despite this difference in bias, the RMSE obtained under contiguity-based W in our model is of a comparable order of magnitude to that reported in LC, suggesting a more favorable bias-variance trade-off in our experiments.

In a word, the results in Table 2 highlight a clear dichotomy in the impact of spatial weight matrix choice. While slope parameters remain robust across alternative spatial structures, the estimation of spatial dependence parameters deteriorates in precision as the spatial weight matrix becomes denser. This finding is consistent with the spatial econometrics literature (Mizruchi & Neuman, 2008), which shows that, for homogeneous SAR models, dense spatial weight matrices tend to yield less reliable estimates of spatial autoregressive parameters, with estimation bias and uncertainty becoming more pronounced as sparsity decreases. In this sense, our evidence directly responds to the concern raised in LC regarding the sensitivity of HSAR specifications to the choice of spatial weights matrices, by demonstrating that such sensitivity primarily manifests in the precision of the heterogeneous dependence parameters rather than in slope estimation. Consequently, when HHSAR models are employed in finite samples, relatively sparse spatial weight matrices would be preferable.

Table 2 Summary statistics of Bayesian estimates for the HHSAR model under alternative spatial weighting matrices

Parameter	Metric	W_{queen}	W_{knn}	W_{Inv}
β_i^1	Bias	0.00055	0.00049	0.00056
	E-SD	0.1152	0.1146	0.1143
	Coverage	0.9474	0.9459	0.9479
	RMSE	0.1168	0.1162	0.1157
β_i^2	Bias	-0.00043	-0.00046	-0.00039
	E-SD	0.09438	0.09423	0.09424
	Coverage	0.9354	0.937	0.9374
	RMSE	0.0987	0.0985	0.0984
ψ_i	Bias	0.0008	0.0019	0.0018
	E-SD	0.2414	0.2777	0.3038
	Coverage	0.9151	0.9258	0.9356
	RMSE	0.2864	0.3158	0.3339

4.3. Bayesian HMC vs. frequentist ML estimation

This subsection reinforces the previous results by directly comparing the performance of the HMC-based Bayesian approach with classical ML estimator. When uninformative priors are assigned to the parameters in an HHSAR model, the Bayesian approach is expected to yield results equivalent to ML estimation[7]. We therefore conduct an experiment to validate our HMC estimation procedure, adopting a design similar to that of LC.

The DGP follows a simpler HHSAR specification with an intercept and a single explanatory variable ($K = 2$). All parameter settings are kept identical to those in Experiment 1, except that β_{i2} is excluded. The W matrix[8] maintains same in Experiment 1.

Under this setup, 500 independent realizations of the dependent variable are generated. The simulation varies the cross-sectional dimension ($N = 25, 100$) and the time dimension ($T = 10, 25, 50, 100, 125$) to assess the performance of the two approaches under small to moderate spatial panels and across both

[7] According to LeSage and Chih (2018), QML and Bayesian approaches based on non-informative priors are theoretically equivalent. Furthermore, Aquaro et al. (2021) show that, under the assumption of normally distributed errors, the QML estimator simplifies to the ML estimator in the HSAR model.

[8] The spatial configuration in our experiment is more complex than the one left- and one right- neighbor structure considered in LC.

short and long-time spans.

Table 3 Summary of performance comparison for the HHSAR model under ML and Bayesian approach

Spatial Units (N)	Metrics	Parameters	HMC			ML estimation		
			10	25	50	10	25	50
25	Bias	α_i	0.00558	-0.00178	0.00161	0.00335	-0.00093	0.00063
		β_i^1	-0.00254	0.00184	-0.00166	-0.00274	0.00199	-0.00173
		ψ_i	-0.00053	-0.00078	0.00068	0.00296	0.00247	0.0015
	RMSE	α_i	0.1808	0.14045	0.11158	0.21655	0.1473	0.11442
		β_i^1	0.17091	0.11698	0.08636	0.21752	0.12721	0.08928
		ψ_i	0.35287	0.32398	0.30263	0.36397	0.31331	0.28902
			HMC			ML estimation		
			50	100	125	50	100	125
100	Bias	ψ_i	0.00045	-0.00077	-0.00026	0.00029	-0.00155	0.00019
		α_i	-0.00037	0.00028	-0.00011	-0.00079	-0.00029	-0.00012
		β_i^1	0.00051	-0.00015	0.00010	-0.00064	-0.00004	-0.00061
	RMSE	ψ_i	0.22275	0.19227	0.15014	0.30489	0.29477	0.29476
		α_i	0.09801	0.07589	0.07218	0.11844	0.09781	0.09513
		β_i^1	0.08401	0.06043	0.05010	0.09041	0.06349	0.05688

Table 3 reports summary statistics for HHSAR parameter estimates under Bayesian HMC and ML estimation across different N and T. Holding T fixed, increasing the number of spatial units generally improves estimation precision for both approaches. Average bias and RMSE decrease consistently for all parameters as N increases under both estimation methods. Increasing the time dimension from $T = 10$ to $T = 25$ leads to substantial gains in accuracy for both intercept and slope parameters, regardless of the estimation approach. These gains are primarily reflected in pronounced reductions in RMSE, while average biases, which is already negligible at $T = 25$, converge even more closely to zero as T increases to 50. Notably, Bayesian HMC exhibits greater robustness than ML estimation in small-sample settings, particularly when $N = 25$ and $T = 10$, consistent with the literature suggesting that Bayesian methods are better suited to small sample contexts (Smid et al., 2020).

The main finding of this analysis is the close correspondence between Bayesian HMC and frequentist ML estimation across most simulation designs. Both methods yield negligible bias and low RMSE for

all parameters. HMC demonstrates an advantage in terms of bias and RMSE in small-sample scenarios with low N and T. As expected, this advantage diminishes as T increases. For moderate sizes of N, HMC continues to deliver marginally lower bias and RMSE than ML.

Overall, results of Experiment 3 indicate that both Bayesian HMC and ML provide accurate and reliable estimation for HHSAR models. The near-equivalence in performance between the two approaches provides researchers with considerable flexibility. Consequently, the choice between Bayesian HMC and ML may be guided by secondary considerations, such as the need for full posterior inference, computational constraints, or the incorporation of prior information, without substantially affecting the quality of point estimates.

4.4. Computational efficiency

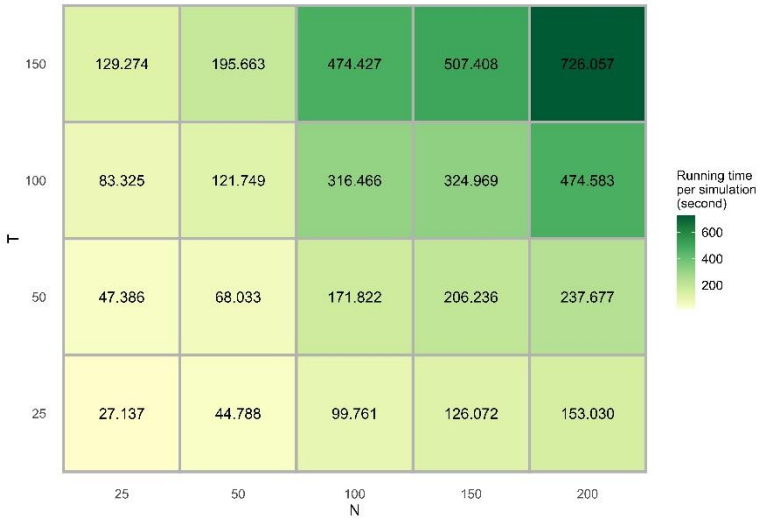


Figure 1 reports computation time (in seconds) for generating 800 HMC draws per chain under different N and T.

We end the simulation analysis by examining the computational efficiency of the Bayesian HMC approach. Computational efficiency is evaluated using total runtime per simulation. Figure 1 presents the time required to generate 1200 HMC draws per chain using 4 chains, based on the same DGP in

Experiment 3.

Two main patterns emerge. First, holding the spatial dimension N fixed, runtime increases monotonically but sublinearly with respect to the temporal dimension T. Relative to a baseline case of $T = 25$, increasing to 50 results in runtime increases of approximately 1.51-1.74 times across all tested values of N ($N \in \{25, 50, 100, 150\}$). Further increases to $T = 150$ and $T = 200$ produce runtime multipliers of 2.57-3.17 and 4.02-4.36, respectively. Second, holding T constant, runtime is also sensitive to changes in the spatial dimension N. Doubling from 25 to 50 increases runtime by a factor of 1.43-1.65. Increasing to 100 (a fourfold increase relative to the baseline) raises runtime by 3.62-3.79 times, while increasing to 150 (a sixfold increase) results in runtime increases of approximately 3.9-4.6 times across all tested values of T. These results correspond to a slightly stronger but still sublinear dependence on N.

When both N and T increase simultaneously, runtime exhibits superlinear but clearly subquadratic growth. For example, scaling from $(N, T) = (25, 25)$ to $(50, 50)$ increases runtime by approximately 2.51 times, substantially less than the fourfold increase implied by purely multiplicative scaling. Similarly, moving to $(100, 100)$ and $(150, 150)$ yields runtime increases of approximately 11.7 and 18.7 times, respectively, compared with the 16-fold and 36-fold increases that would be expected under quadratic scaling.

To summarize, these results suggest that, over the examined settings, the computational cost of our Bayesian HMC implementation exhibits similarly sensitive to increases in both the spatial and temporal dimensions. This differs somewhat from the timing behavior documented in LC, where runtime under their M-H approach responds more strongly to increases in N. Although our Bayesian HMC

implementation remains more computationally demanding than their fine-tuned M-H algorithm, the observed runtimes we observed are generally within practical and acceptable limits for most applied research contexts.

5. An empirical illustration

5.1 Background and motivation

As an empirical illustration, we examine the relationship between economic growth and renewable energy consumption in a panel of 25 European countries over the period 1991-2015, following the data construction in Chica-Olmo et al. (2020). Renewable energy development has attracted sustained attention from both researchers and policymakers, as modern economies face the dual challenge of achieving energy sustainability while maintaining economic competitiveness and affordability.

A growing literature documents both the economic and environmental benefits of renewable energy adoption. Increasing the share of renewable energy has been shown to enhance productivity and economic efficiency, reduce reliance on fossil fuels, and strengthen energy security through diversification of supply sources (Inglesi-Lotz, 2016). Renewable energy deployment is also widely recognized as an important instrument for mitigating climate change and lowering greenhouse gas emissions (Bhuiyan et al., 2022). Beyond its direct economic effects, renewable energy development exhibits a clear spatial dimension: policies, investment decisions, and outcomes are not spatially independent but are shaped by cross-country interactions. Empirical evidence documents significant spatial dependence in renewable energy dynamics, indicating that developments in one country systematically influence neighboring countries (Ma et al., 2023), thereby motivating the explicit incorporation of spatial dependence in empirical modeling.

In the European context, spatial spillovers are expected to be important. We distinguish between spill-in

effects, whereby a country's economic performance is influenced by developments in neighboring countries, and spill-out effects, whereby domestic conditions affect economic outcomes elsewhere. Geographic proximity facilitates policy imitation and competitive pressures. Institutionally mediated spillovers within the EU further strengthen spatial dependence. Common regulatory targets, market-based instruments such as the EU Emissions Trading System, and cross-border energy infrastructure create structured channels for policy diffusion, generating spill-out effects from frontrunner countries and spill-in effects for lagging members through regulatory alignment, funding, and technology transfer.

5.2 Homogeneous SAR panel model estimates

We begin with a homogeneous SAR panel model as a benchmark. GDP per capita is used as a proxy for economic growth, and the empirical specification follows a Cobb-Douglas production framework. Output is modeled as a function of renewable energy consumption per capita (REC), capital stock measured by gross fixed capital formation per capita (GFCF), and the labor force (LABOR), consistent with Inglesi-Lotz (2016). We employ the weights matrix using a k-nn scheme with $k = 6$ for this application.

Table 4 Homogeneous SAR panel model with fixed effects estimates.

Variable	Estimates	Direct effects	Indirect effects	Total effects
REC	0.005	0.006	0.004	0.010
GFCF	0.339***	0.347***	0.217***	0.564***
LABOR	0.283***	0.244***	0.152***	0.396***
Homogeneous spatial autoregressive parameter ρ	0.396***	-	-	-

Note: The homogeneous SAR panel model with fixed effects has the following specification: $GDP_{it} = \alpha + \rho GDP_{jt} + \beta^{(1)}REC + \beta^{(2)}GFCF + \beta^{(3)}LABOR + \upsilon_i + \nu_t + \varepsilon_{it}$.

Table 4 reports results from a homogeneous SAR panel model with fixed effects. The estimates indicate moderate, positive, and statistically significant spatial dependence in GDP per capita across countries ($\rho = 0.396^{***}$). With the exception of REC, both GFCF and LABOR exhibit positive and significant coefficients, consistent with theoretical expectations. The decomposition into direct, indirect, and total effects shows consistent signs across all explanatory variables.

5.3 HHSAR panel model estimates

We now turn to the HHSAR panel model, $GDP_{it} = \alpha_i + \psi_i \sum_{j=1}^{25} w_{ij}\, GDP_{jt} + \beta_i^{(1)} REC + \beta_i^{(2)} GFCF + \beta_i^{(3)} LABOR + \varepsilon_{it}$. We first present the country-specific estimates of the spatial autoregressive parameters, followed by the corresponding direct, spill-in, and spill-out effects.

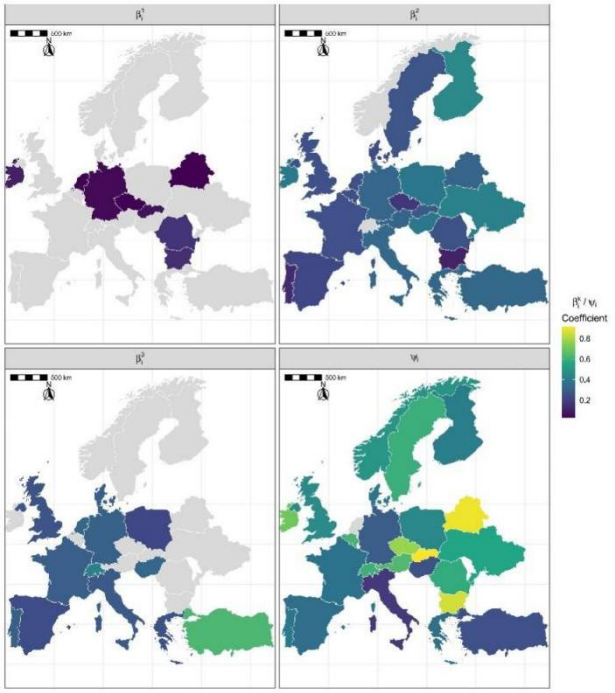


Figure 2. Spatial distribution of posterior means for random slopes and spatial autoregressive parameters. Panels (a)-(c) report country-specific posterior means of the random slope coefficients β_i^k for the three regressors, and panel (d) reports the posterior mean of the heterogeneous spatial autoregressive parameter ψ_i. Only estimates with 95% CIs excluding zero are mapped; non-significant countries are shaded in grey.

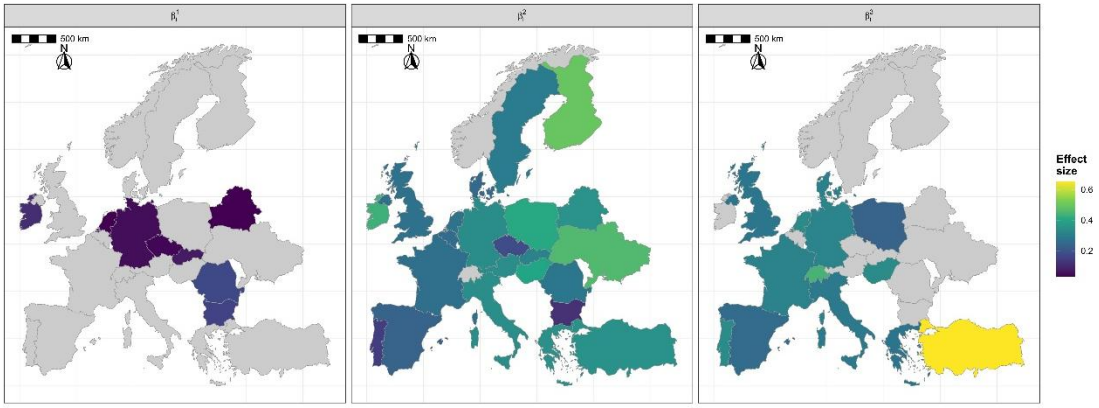
Figure 2 (lower-right panel) illustrates the spatial distribution of the posterior mean estimates of the heterogeneous spatial dependence parameter (Posterior means and 95% CIs are reported in Table 1 of

Appendix A.). Allowing ψ_i to vary across countries provides a flexible representation of cross-border spillovers and reveals substantial heterogeneity in their strength. For most countries, estimated ψ_i are positive and statistically significant, indicating that domestic economic growth is systematically influenced by growth dynamics in neighboring countries. However, the magnitude of this dependence varies markedly across space.

Several Central and Eastern European countries, such as Slovakia, Poland, Romania, Belarus, and Ukraine, exhibit relatively high ψ_i estimates, suggesting strong sensitivity to country-level spillovers. This pattern is consistent with ongoing convergence processes, tight cross-border production linkages. In contrast, core Western European economies, including Germany, France, and the United Kingdom, display more moderate but still significant ψ_i estimates, reflecting diversified economic structures or broader sets of trading partners that may reduce reliance on immediate neighbors.

5.4 Direct effects estimates

We next assess the direct effect which measures the impact of changes in a country's own explanatory variables on its domestic economic performance. Figure 3 maps the spatial distribution of posterior mean direct effects for REC, GFCF, and LABOR[9].



[9] Tables 2, 3 and 4 of Appendix A report posterior mean estimates and 95% CIs for direct effects, spill-in and spill-out effects across the three explanatory variables.

Figure 3. Panels show country-specific direct effects for $\beta_i^{(1)}$, $\beta_i^{(2)}$, and $\beta_i^{(3)}$, with non-significant effects omitted (grey). Bright colors represent larger direct effects.

For REC, statistically significant positive direct effects are observed primarily among a subset of EU member states in Central and Eastern Europe, such as Bulgaria, Hungary, and Greece, as well as a small number of non-EU countries. This suggests that renewable energy consumption contributes more strongly to domestic growth in economies for lagging countries in EU or for countries with ongoing energy system reconfiguration. By contrast, many Western and Northern EU countries exhibit statistically insignificant direct effects of REC.

Direct effects associated with capital accumulation are more widespread and generally significant across both EU and non-EU countries, consistent with the central role of capital accumulation in economic growth. By comparison, direct effects of the labor force exhibit greater heterogeneity, with significant effects observed in most Western EU countries.

Overall, the pronounced heterogeneity in direct effects across Europe highlights the limitations of homogeneous models and motivates a more detailed examination of cross-border spillovers.

5.5 Spill-in effects estimates

We next consider spill-in effects, which capture the cumulative impact of changes in neighboring countries' explanatory variables on a given country's economic performance.

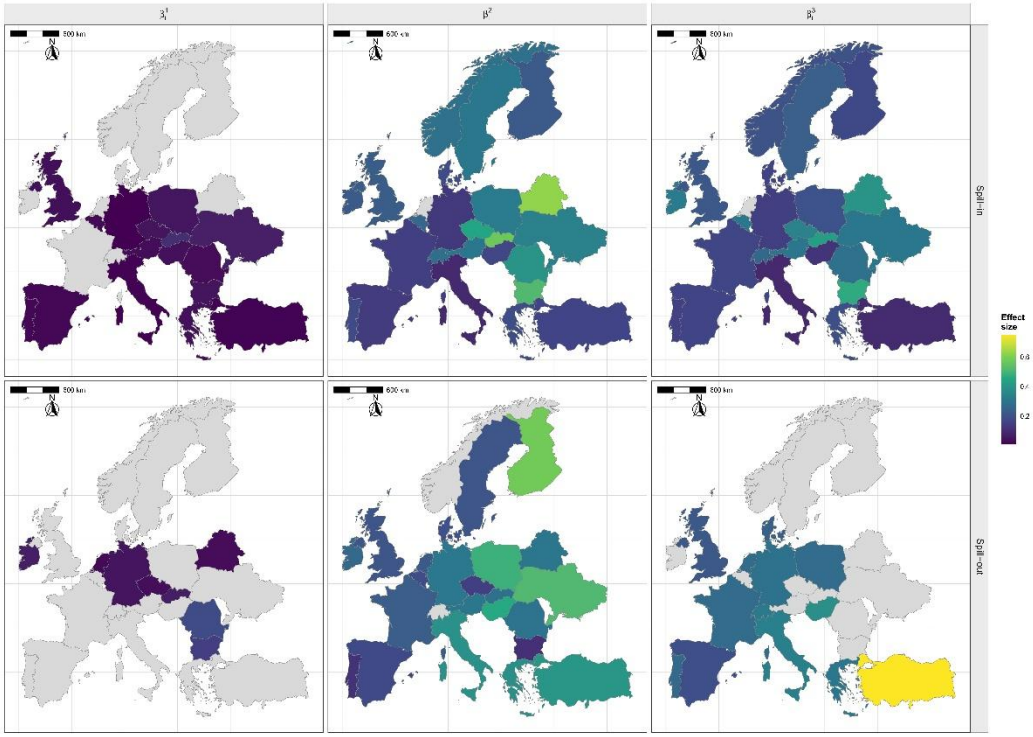


Figure 4. Spatial patterns of posterior mean spill-in (top row) and spill-out (bottom row) effects. Spill-in effects summarize incoming spatial impacts from neighboring countries, whereas spill-out effects capture outgoing impacts transmitted to neighbors. Only statistically significant effects are displayed.

Figure 4 (upper panels) show statistically significant spill-in effects are relatively selective. For REC, significant spill-in effects are observed for only a limited number of countries, most of which are EU member states located in Central and Eastern Europe. This finding suggests that renewable energy consumption in neighboring countries primarily influences domestic growth in economies that are still converging or undergoing energy system reconfiguration.

Spill-in effects associated with GFCF are more prevalent, particularly among EU countries. Several EU member states display positive and statistically significant spill-in effects from neighboring capital accumulation. In contrast, peripheral and non-EU countries generally exhibit weaker spill-in effects. Spill-in effects related to the labor force exhibit patterns similar to those observed for GFCF.

5.6 Spill-out effects estimates

The lower panels of Figure 4 reveal pronounced asymmetry in spill-out effects. For REC, only a small

number of countries exert statistically significant spill-out effects, and these are predominantly EU member states.

Spill-out effects associated with GFCF are more widespread. Several economically influential EU member states exert significant positive spill-out effects through capital accumulation. Spill-out effects for the LABOR are limited and sporadic, yet a core-periphery pattern emerges. When significant spill-out effects occur, they are again concentrated in core economies such as France and Germany, reinforcing the notion that labor-related growth effects remain largely domestic. Importantly, spill-out effects do not necessarily coincide with spill-in effects, some EU countries act primarily as sources of spillovers, while others are mainly recipients. This asymmetry highlights the uneven spatial transmission of economic forces across Europe and underscores the importance of distinguishing between inward and outward spillover channels.

6. Conclusions

This study contributes to the growing literature on HSAR panel models in several important ways. First, we extend existing HSAR specifications by adopting a hierarchical prior with partial pooling, thereby relaxing the large-T requirements commonly imposed in the conventional HSAR specification such as those in Aquaro et al. (2015). This extension enables reliable inference in panels with small to moderate time dimensions, which are typical in empirical regional and spatial economic applications.

Second, we demonstrate the HMC-based integrated Bayesian approach as an effective tool for estimating hHSAR parameters. Monte Carlo simulations demonstrate that our HHSAR specification performs well across a wide range of sample configurations via the integrated Bayesian approach, including scenarios with $N > T$, $N = T$ and $N < T$. Estimation accuracy improves with sample size, and the approach consistently outperforms ML estimation in small-sample settings. While slope parameters are generally

robust to alternative spatial weight matrices, the precision of spatial dependence parameters declines as spatial connectivity becomes denser. Overall, the proposed approach achieves a favorable balance between statistical performance and computational cost, with runtimes suitable for most applied studies.

An empirical application to renewable energy consumption and economic growth in European countries illustrates the practical relevance of the HHSAR model. Whereas homogeneous SAR models capture only average spatial interaction, the HHSAR model uncovers substantial cross-country heterogeneity in both direct and spillover effects. Spill-in and spill-out effects are asymmetric, reflecting core-periphery structures and institutional integration.

Several promising avenues for future research remain. One direction is the development of clustered HSAR models, in which regions are endogenously grouped into a limited number of clusters sharing common spatial dependence parameters. Such an approach would provide an intermediate structure between homogeneous and fully heterogeneous specifications, improving interpretability while preserving flexibility. Another challenge concerns the construction of spatial weight matrices. Although our results suggest relative robustness to alternative specifications, doubly stochastic normalization could be problematic in small samples, motivating further work on regularized or alternative normalization schemes. Finally, by replacing HMC sampling with optimization-based posterior approximation, variational inference could substantially reduce computational burden in models featuring region-specific spatial autoregressive parameters and large spatial weight matrices.

Reference

Aquaro, M., Bailey, N., & Pesaran, M. H. (2015). Quasi maximum likelihood estimation of spatial models with heterogeneous coefficients. *USC-INET Research Paper*, (15–17).

Aquaro, M., Bailey, N., & Pesaran, M. H. (2021). Estimation and inference for spatial models with heterogeneous coefficients: An application to US house prices. *Journal of Applied Econometrics*, *36*(1), 18–44. https://doi.org/https://doi.org/10.1002/jae.2792

Autant-Bernard, C., & LeSage, J. (2019). A heterogeneous coefficient approach to the knowledge production function. *Spatial Economic Analysis*, *14*(2), 196–218.

Avron, H., & Toledo, S. (2011). Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *Journal of the ACM (JACM)*, *58*(2), 1–34.

Betancourt, M. (2019). The convergence of Markov chain Monte Carlo methods: from the Metropolis method to Hamiltonian Monte Carlo. *Annalen Der Physik*, *531*(3), 1700214.

Betancourt, M., Byrne, S., Livingstone, S., & Girolami, M. (2017). The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli*, *23*(4A), 2257–2298. https://doi.org/10.3150/16-BEJ810

Bhuiyan, M. A., Zhang, Q., Khare, V., Mikhaylov, A., Pinter, G., & Huang, X. (2022). Renewable Energy Consumption and Economic Growth Nexus—A Systematic Literature Review. *Frontiers in Environmental Science*, *Volume 10-2022*.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Chen, D.-G., & Fraser, M. W. (2017). A Bayesian Approach to Sample Size Estimation and the Decision to Continue Program Development in Intervention Research. *Journal of the Society for Social Work and Research*, *8*(3), 457–470. https://doi.org/10.1086/693433

Chib, S. (2008). Panel data modeling and inference: a Bayesian primer. In *The econometrics of panel data: Fundamentals and recent developments in theory and practice* (pp. 479–515). Springer.

Chica-Olmo, J., Sari-Hassoun, S., & Moya-Fernández, P. (2020). Spatial relationship between economic growth and renewable energy consumption in 26 European countries. *Energy Economics*, *92*, 104962.

Davis, C., & Vaddady, P. (2025). Within-chain parallelization—Giving Stan Jet Fuel for population modeling in pharmacometrics. *CPT: Pharmacometrics & Systems Pharmacology*, *14*(1), 52–67.

Gao, J., Liu, B., Ji, W., & Huang, H. (2024). A systematic literature survey of sparse matrix-vector multiplication. *ArXiv Preprint ArXiv:2404.06047*.

Gelman, A. (2006). *Prior distributions for variance parameters in hierarchical models (comment on article*

by Browne and Draper).

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 555.

Han, I., Malioutov, D., & Shin, J. (2015). Large-scale log-determinant computation through stochastic Chebyshev expansions. *International Conference on Machine Learning*, 908–917.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, *15*(1), 1593–1623. https://doi.org/10.48550/arXiv.1111.4246

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, *5*(9).

Hsiao, C., & Pesaran, M. H. (2008). Random Coefficient Models. In L. Mátyás & P. Sevestre (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (pp. 185–213). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75892-1_6

Hutchinson, M. F. (1989). A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, *18*(3), 1059–1076.

Inglesi-Lotz, R. (2016). The impact of renewable energy consumption to economic growth: A panel data application. *Energy Economics*, *53*, 58–63.

Knight, P. A. (2008). The Sinkhorn–Knopp algorithm: convergence and applications. *SIAM Journal on Matrix Analysis and Applications*, *30*(1), 261–275.

LeSage, J., & Chih, Y.-Y. (2016). Interpreting heterogeneous coefficient spatial autoregressive panel models. *Economics Letters*, *142*, 1–5.

LeSage, J., & Chih, Y.-Y. (2018). A Matrix Exponential Spatial Panel Model with Heterogeneous Coefficients. *Geographical Analysis*, *50*(4), 422–453. https://doi.org/https://doi.org/10.1111/gean.12152

LeSage, J. P., & Chih, Y.-Y. (2018). A Bayesian spatial panel model with heterogeneous coefficients. *Regional Science and Urban Economics*, *72*, 58–73.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

Ma, L., Wang, Q., Shi, D., & Shao, Q. (2023). Spatiotemporal patterns and determinants of renewable energy innovation: Evidence from a province-level analysis in China. *Humanities and Social Sciences Communications*, *10*(1), 357. https://doi.org/10.1057/s41599-023-01848-y

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

Mizruchi, M. S., & Neuman, E. J. (2008). The effect of density on the level of bias in the network autocorrelation model. *Social Networks*, *30*(3), 190–200. https://doi.org/https://doi.org/10.1016/j.socnet.2008.02.002

Mukherjee, A., & Vats, D. (2026). Hamiltonian Monte Carlo for (Physics) Dummies. *ArXiv Preprint ArXiv:2601.01422*.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, *2*(11), 2.

Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2020). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 131–161.

Tan, C., & Elhorst, J. P. (2023). The Spatial Autoregressive Panel Data Model with Spatial Moving Average Errors. *Geographical Analysis*. https://doi.org/10.1111/gean.12369

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, *16*(2), 667–718. https://doi.org/10.1214/20-BA1221

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendixv1.pdf