Spatial Statistics 70 (2025) 100931

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Model averaging for spatial autoregressive panel data models

Aibing Ji, Jingxuan Li, Qingqing Li *

College of Mathematics and Information Science, Hebei University, Baoding, 071002, China

A R T I C L E I N F O

Keywords:
Spatial autoregressive panel data models
Model averaging
Mallows-type criterion
Spatiotemporal data analysis

A B S T R A C T

The spatial autoregressive panel data models are widely employed in regional economics to capture spatial dependencies, but conventional specifications rely on a single spatial weight matrix, heightening the risk of model misspecification. Current research lacks systematic model averaging methods for integrating multiple weight matrices and addressing spatial effect uncertainty. This study proposes a novel model averaging framework for spatial autoregressive panel data models with fixed effects, extending model averaging methodology to the spatial panel context and enabling flexible combinations of multiple weight matrices for both dependent variables and error terms. An adaptive Mallows-type criterion is developed, dynamically adjusting to the presence or absence of spatial effects, with its asymptotic optimality established. Monte Carlo simulations confirm robustness across scenarios with no, single, or mixed spatial dependencies. An empirical application to Chinese provincial housing prices identifies economic adjacency as the key spatial dependence driver, validating the method's predictive accuracy and policy utility for spatiotemporal data analysis.

1. Introduction

Spatial econometrics provides powerful tools for modeling spatial dependencies in data, with applications in regional economics, environmental science, and urban studies (Anselin, 1988). A cornerstone of this field is the spatial autoregressive models with autoregressive disturbances (SARAR) (Kelejian and Prucha, 2010), which capture spatial interactions in both dependent variables and error terms. However, traditional SARAR models typically rely on a single spatial weight matrix to define spatial relationships, such as geographical proximity or economic connections (LeSage and Pace, 2009). This reliance poses a significant risk of model misspecification, as real-world spatial dependencies are often complex and may arise from multiple sources. For instance, in regional housing price analysis, spatial effects may stem from both geographical adjacency and economic linkages, and an incorrectly specified weight matrix can lead to biased estimates and unreliable predictions. Therefore, effectively integrating information from different spatial structures to enhance model robustness and prediction accuracy has become a critical challenge in the field of spatial econometrics.

To address model uncertainty, econometricians have developed selection criteria like the Akaike information criterion (AIC) (Akaike, 1973) and Bayesian information criterion (BIC) (Schwarz, 1978). Additionally, the development of specification tests like the spatial J-test (Kelejian, 2008) and its extensions (Kelejian and Piras, 2011) has provided important tools for detecting model misspecification in spatial contexts. Specification tests such as the J-test for spatial autoregressive binary models can help in validating the choice of spatial structure (Piras and Sarrias, 2025). However, these methods select a single "best" model but fail to account for uncertainty across multiple plausible specifications.

* Corresponding author.
E-mail address: liqingqing@stumail.hbu.edu.cn (Q. Li).

https://doi.org/10.1016/j.spasta.2025.100931
Received 13 June 2025; Received in revised form 7 August 2025; Accepted 9 September 2025
Available online 23 September 2025

Model averaging (MA) provides a principled framework to mitigate model uncertainty by combining estimates across multiple specifications, weighting them according to predictive performance (Hansen, 2007). Within spatial econometrics, LeSage and Parent (2007) pioneered Bayesian model averaging (BMA) for spatial models, while Zhang and Yu (2018) developed frequentist MA methods for selecting and averaging weight matrices in cross-sectional spatial autoregressive (SAR) models. Nguyen and Veraart (2017) introduced a novel approach using volatility modulated moving averages to model spatial heteroskedasticity, where variances and covariances vary stochastically across space, offering robust applications in fields like environmental science. However, these approaches face challenges when applied to SARAR panel models: (i) BMA methods are computationally intensive, especially for large panels; (ii) existing frequentist MA methods (Zhang and Yu, 2018) focus primarily on cross-sectional SAR models and do not address the complexities of SARAR structures with panel fixed effects; (iii) no method systematically handles the combinatorial selection of multiple distinct weight matrices ($\mathbf{W}_s$ for the dependent variable and $\mathbf{M}_h$ for the errors) within the SARAR panel framework; and (iv) current criteria lack mechanisms to adaptively handle degenerate cases (e.g., $\lambda = 0$ or $\rho = 0$), potentially leading to inconsistent model selection when spatial effects are absent.

Recent advances in spatial model averaging and uncertainty have expanded the scope of spatial econometric methods. For instance, Ben Abdallah et al. (2025) explore stock return co-movements using a convex combination of spatial weight matrices, demonstrating that financial-distance measures outperform geographical ones. In traffic analysis, Similarly, Chen et al. (2025) apply functional SAR model averaging with a Mallows criterion to estimate urban traffic density, constructing macroscopic fundamental diagrams for Guiyang City. Additionally, Song and Cibin (2024) optimize spatial weight matrices using graph-based shortest path algorithms, improving crime event predictions in New York City, while Su et al. (2023) introduce a Bayesian spatiotemporal autoregressive (STAR) model with spatiotemporal autoregressive conditional heteroscedasticity (STARCH) errors, employing Bayesian model averaging to capture dependence in mean and variance simultaneously. While these studies advance spatial modeling, they focus on cross-sectional, functional, or Bayesian contexts and do not address SARAR panel models with multiple weight matrices and fixed effects.

Recent advances in time series model averaging (Liao et al., 2021; Lin and Liu, 2025; Zou et al., 2025), though not directly addressing spatial dependence, provide valuable insights for managing model uncertainty and complexity. Similarly, work on spatial panel estimation (Lee and Yu, 2010; Jin and Lee, 2020; Tian et al., 2025) and interval-valued models (Li et al., 2025) offers relevant methodologies but does not resolve the fundamental gap.

This study addresses these gaps by introducing a novel model averaging framework for SARAR panel models, tackling uncertainty in spatial dependence structures through the integration of multiple spatial weight matrices. The Mallows-type model averaging approach (Hansen, 2007) is extended to the spatial panel context, incorporating an adaptive criterion that dynamically adjusts to the presence or absence of spatial effects. Orthogonal transformations are employed to eliminate individual fixed effects while preserving error structures, ensuring computational efficiency. Monte Carlo simulations confirm the framework's robustness across scenarios with no, single, or mixed spatial dependencies, while an empirical application to Chinese provincial housing prices (2013–2022) highlights its practical utility in identifying economic adjacency as a primary driver of spatial dependence.

Our contributions advance the field of spatial econometrics in four significant ways:

- Extend the Mallows-type model averaging method from spatial cross-sectional to spatial panel models;
- Provide a systematic treatment of model averaging for SARAR panel models with multiple spatial weight matrices;
- Develop an adaptive model selection criterion that adjusts to the presence or absence of spatial effects;
- Offer a computationally efficient implementation for practical empirical applications.

These advancements enhance the robustness and predictive accuracy of spatial econometric models, offering a versatile tool for researchers and policymakers analyzing spatiotemporal data.

The rest of the paper is organized as follows. Section 2 establishes the spatial panel SARAR model framework and derives the quasi-maximum likelihood (QML) estimation procedure. Section 3 proposes the Mallows-type model averaging framework with adaptive penalty mechanisms and its asymptotic properties are proved. Section 4 systematically examines three degenerate model specifications. Monte Carlo simulation designs and empirical results are presented in Sections 5 and 6, respectively. We conclude with policy implications and future research directions in Section 7. Technical proofs and additional derivations are provided in Appendix A.

2. Model and estimation

We consider a spatial autoregressive panel model with fixed effects and SAR disturbances

$$Y_{nt}^* = \lambda W_n Y_{nt}^* + X_{nt}^*\beta + c_{n0} + U_{nt}^*, \quad U_{nt}^* = \rho M_n U_{nt}^* + V_{nt}^*, \quad t = 1, \dots, T,$$

where $Y_{nt}^* = (y_{1t}^*, \dots, y_{nt}^*)'$ and $V_{nt}^* = (v_{1t}^*, \dots, v_{nt}^*)'$ are $n \times 1$ vectors, with $v_{it}^* \sim$ i.i.d.$(0, \sigma_0^2)$ across i and t. The matrix X_{nt}^* (of dimension $n \times k$) contains nonstochastic time varying regressors, and c_{n0} is an $n \times 1$ vector of fixed effects. The $n \times n$ nonstochastic spatial weights matrix W_n captures cross-sectional spatial dependence among the units in y_{it}^*, M_n is a $n \times n$ spatial weights matrix related to the error. The detailed explanations of the symbols used can be found in Table 10 of Appendix B.

Consider a panel dataset containing n cross-sectional units over T time periods,

$$\mathbf{Y}^* = \lambda(\mathbf{I}_T \otimes \mathbf{W}_n)\mathbf{Y}^* + \mathbf{X}^*\boldsymbol{\beta} + \mathbf{I}_T \otimes \mathbf{c}_{n0} + \mathbf{U}^*, \quad \mathbf{U}^* = \rho(\mathbf{I}_T \otimes \mathbf{M}_n)\mathbf{U}^* + \mathbf{V}^*, \tag{1}$$

where stacked dependent variable $\mathbf{Y}^* = \left(Y_{n1}^{*\prime},\ Y_{n2}^{*\prime},\ \cdots\ Y_{nT}^{*\prime}\right)'$ and stacked error term $\mathbf{V}^* = \left(V_{n1}^{*\prime}, V_{n2}^{*\prime},\ \cdots\ V_{nT}^{*\prime}\right)'$ are $NT \times 1$ vectors. Design matrix of explanatory variables $\mathbf{X}^*$ is an $NT \times K$ vector as follow:

$$\mathbf{X}^* = \begin{bmatrix} x_{11,1}^* & x_{11,2}^* & \cdots & x_{11,K}^* \\ x_{12,1}^* & x_{12,2}^* & \cdots & x_{12,K}^* \\ \vdots & \vdots & \ddots & \vdots \\ x_{21,1}^* & x_{21,2}^* & \cdots & x_{21,K}^* \\ \vdots & \vdots & \ddots & \vdots \\ x_{nT,1}^* & x_{nT,2}^* & \cdots & x_{nT,K}^* \end{bmatrix}$$

To ensure consistent estimation of the variance parameter (see the supplementary materials), we employ orthogonal transformations to eliminate individual effects. An orthogonal transformation matrix $\mathbf{F}_{T,T-1}$ of dimension $T\times(T-1)$ is utilized while preserving the independence structure of error terms. This matrix satisfies the following key properties:

$$\mathbf{F}_{T,T-1}'\mathbf{F}_{T,T-1} = \mathbf{I}_{T-1}, \quad \mathbf{F}_{T,T-1}'\mathbf{e}_T = \mathbf{0},$$

thus it has

$$\mathbf{F}_{T,T-1}\mathbf{F}_{T,T-1}' = \mathbf{I}_T - \frac{1}{T}(\mathbf{e}_T\mathbf{e}_T'),$$

where $\mathbf{I}_{T-1}$ is the identity matrix of dimension $(T-1)\times(T-1)$, and $\mathbf{e}_T$ is a $T \times 1$ vector of ones. The first condition ensures the orthogonality of $\mathbf{F}_{T,T-1}$, while the second condition guarantees that the transformation removes the time-invariant individual fixed effects by projecting the data onto a subspace orthogonal to the vector of ones.

The transformation operator $\mathbf{Q} = \mathbf{F}_{T,T-1}' \otimes \mathbf{I}_n$ is constructed as the Kronecker product of $\mathbf{F}_{T,T-1}$ and the $n \times n$ identity matrix $\mathbf{I}_n$. This operator is applied to the stacked panel data to perform the transformation. Specifically, for the dependent variable $\mathbf{Y}^*$, the explanatory variables $\mathbf{X}^*$, and the error term $\mathbf{V}^*$, the transformed versions are obtained as follows:

$$\mathbf{Y} = \mathbf{QY}^* = \left(Y_{n1}',\ Y_{n2}',\ \cdots\ Y_{n(T-1)}'\right)'; \quad \mathbf{X} = \mathbf{QX}^* = \left(X_{n1}',\ X_{n2}',\ \cdots\ X_{n(T-1)}'\right)';$$

$$\mathbf{U} = \mathbf{QU}^* = \left(U_{n1}',\ U_{n2}',\ \cdots\ U_{n(T-1)}'\right)'; \quad \mathbf{V} = \mathbf{QV}^* = \left(V_{n1}',\ V_{n2}',\ \cdots\ \mathbf{V}_{n(T-1)}'\right)'.$$

The transformed SARAR panel model takes the following form:

$$\mathbf{Y} = \lambda(\mathbf{I}_T \otimes \mathbf{W}_n)\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{I}_T \otimes \mathbf{c}_{n0} + \mathbf{U}, \quad \mathbf{U} = \rho(\mathbf{I}_T \otimes \mathbf{M}_n)\mathbf{U} + \mathbf{V}.$$

The reduced form of the SARAR panel model is

$$\mathbf{Y} = (\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))^{-1}(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho))^{-1}\mathbf{V}, \tag{2}$$

where the spatial transformation matrices are defined as $\mathbf{S}_n(\lambda) = \mathbf{I}_n - \lambda\mathbf{W}_n$ and $\mathbf{R}_n(\rho) = \mathbf{I}_n - \rho\mathbf{M}_n$, with $\mathbf{S}_n = \mathbf{S}_n(\lambda)$ and $\mathbf{R}_n = \mathbf{R}_n(\rho)$ corresponding to the true parameter values λ and ρ, respectively. The model comprises a nonstochastic component, $(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))^{-1}\mathbf{X}\boldsymbol{\beta}$, and a stochastic component, $(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))^{-1}(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho))^{-1}\mathbf{V}$. The conditional expectation of the dependent variable is expressed as:

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{Y}) = (\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))^{-1}\mathbf{X}\boldsymbol{\beta}, \tag{3}$$

and the variance–covariance matrix of $\mathbf{Y}$ is given by $\boldsymbol{\Omega} = \sigma_0^2\left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda)\mathbf{R}_n(\rho)\right)^{-1}\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho)'\mathbf{S}_n(\lambda)'\right)^{-1}$.

To account for uncertainty in spatial dependence, we consider sets of candidate spatial weight matrices $\mathcal{W} = \{\mathbf{W}_1, \ldots, \mathbf{W}_S\}$ for the dependent variable and $\mathcal{M} = \{\mathbf{M}_1, \ldots, \mathbf{M}_H\}$ for the error term, yielding $S \times H$ model specifications. For each candidate model (s, h), the QML estimator (Lee and Yu, 2010) is derived from the log-likelihood:

$$\ln L(\boldsymbol{\theta}) = -\frac{n(T-1)}{2}\ln(2\pi\sigma_0^2) + \ln|(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda_{s,h}))| + \ln|\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho_{s,h})| - \frac{1}{2\sigma_0^2}\mathbf{V}'(\boldsymbol{\theta})\mathbf{V}(\boldsymbol{\theta}), \tag{4}$$

where the transformed error vector is $\mathbf{V}(\boldsymbol{\theta}) = \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho_{s,h})\right)\left[\left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda_{s,h})\right)\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right]$.

The parameter estimates for the sth candidate model $\mathbf{W}_s$ and the hth model $\mathbf{M}_h$ are

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{s,h} = {} & \left[\mathbf{X}'\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})'\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right)\mathbf{X}\right]^{-1} \\ & \times \mathbf{X}'\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})'\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right)\mathbf{Y},\end{aligned} \tag{5}$$

and

$$\begin{aligned}\hat{\sigma}_{s,h}^2 = {} & \frac{1}{n(T-1)}\left[\left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right)\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{s,h}\right]' \\ & \times \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})'\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right)\left[\left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right)\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{s,h}\right].\end{aligned}$$

The resulting estimator for the expected value $\boldsymbol{\mu}$ is:

$$\hat{\boldsymbol{\mu}}_{s,h} = \left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right)^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}_{s,h}, \tag{6}$$

which can be equivalently expressed as

$$\hat{\boldsymbol{\mu}}_{s,h} = \left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda_{s,h})\right)^{-1} \mathbf{X} \left[\mathbf{X}' \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})'\right) \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right) \mathbf{X}\right]^{-1} \times \mathbf{X}' \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})'\right) \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right) \left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right) \mathbf{Y}. \tag{7}$$

The projection matrix $\tilde{\mathbf{P}}_{s,h}$ in this context is defined as

$$\tilde{\mathbf{P}}_{s,h} = \left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda_{s,h})\right)^{-1} \mathbf{X} \left[\mathbf{X}' \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})'\right) \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right) \mathbf{X}\right]^{-1} \times \mathbf{X}' \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})'\right) \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right) \left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right), \tag{8}$$

where the transformed error vector is $\mathbf{V}(\theta) = (\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho_{s,h})) \left[(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda_{s,h}))\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right]$.

For analytical purposes, we obtain the following derivatives:

$$\frac{\partial \hat{\rho}_{s,h}}{\partial \mathbf{Y}^\top} = -\left(\frac{\partial^2 \ln L(\theta)}{\partial \rho^2}\right)^{-1} \frac{2}{\sigma^2} (\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))' (\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho))' (\mathbf{I}_{T-1} \otimes \mathbf{M}_n) \mathbf{Z}, \tag{9}$$

$$\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\rho}_{s,h}} = \left(\mathbf{A} \frac{\partial \mathbf{B}}{\partial \hat{\rho}_{s,h}} \mathbf{C} + \mathbf{A}\mathbf{B} \frac{\partial \mathbf{C}}{\partial \hat{\rho}_{s,h}}\right), \tag{10}$$

$$\frac{\partial \hat{\lambda}_{s,h}}{\partial \mathbf{Y}^\top} = -\left(\frac{\partial^2 \ln L(\theta)}{\partial \lambda^2}\right)^{-1} \frac{2}{\sigma^2} (\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho))(\mathbf{I}_{T-1} \otimes \mathbf{W}_n) \mathbf{V}(\theta), \tag{11}$$

$$\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\lambda}_{s,h}} = \left(\frac{\partial \mathbf{A}}{\partial \hat{\lambda}_{s,h}} \mathbf{B}\mathbf{C} + \mathbf{A}\mathbf{B} \frac{\partial \mathbf{C}}{\partial \hat{\lambda}_{s,h}}\right). \tag{12}$$

The derivation of Eqs. (9)–(12), along with the definitions of **A**, **B**, and **C** are detailed in Appendix A. For Eqs. (9) and (11), the primary approach involves differentiating both sides of the score equation, following Greven and Kneib (2010). These equations are instrumental in developing the model selection criterion discussed in the subsequent section.

3. Model averaging framework

3.1. Model averaging procedure

We consider S candidate spatial weight matrices $\mathbf{W}_s$ ($s = 1, \ldots, S$) for the dependent variable and H matrices $\mathbf{M}_h$ ($h = 1, \ldots, H$) for the error terms, where $\mathbf{W}_1$ and $\mathbf{M}_1$ represent null matrices (no spatial effects). We allow the dimensions S and H to grow with the sample size $n(T-1)$, ensuring the asymptotic validity of our approach as the number of observations increases. This framework generates $S \times H$ possible model specifications, each indexed by a unique pair (s, h). Let $\mathbf{w} = (w_{1,1}, \ldots, w_{S,H})'$ belonging to the set $\mathcal{H} = \left\{\mathbf{w} \in [0,1]^{SH} : \sum_{s=1}^{S} \sum_{h=1}^{H} w_{s,h} = 1\right\}$, and define $\tilde{\mathbf{P}}(\mathbf{w}) = \sum_{s=1}^{S} \sum_{h=1}^{H} w_{s,h} \tilde{\mathbf{P}}_{s,h}$. The model average estimator of $\boldsymbol{\mu}$ would then be

$$\hat{\boldsymbol{\mu}}(\mathbf{w}) = \tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} = \sum_{s=1}^{S} \sum_{h=1}^{H} w_{s,h} \hat{\boldsymbol{\mu}}_{s,h}. \tag{13}$$

This paper addresses the challenge of selecting an optimal spatial weight matrix by adopting a model averaging approach that integrates multiple candidate models to minimize the risk of suboptimal selection. By combining predictions from diverse models, model averaging enhances robustness against model uncertainty, overcoming the limitations of relying on a single specification. Drawing on the literature on model selection and averaging (Hansen, 2007; Zhang and Yu, 2018), a criterion function is developed, tailored to the spatial econometric framework. The objective is to minimize the squared loss, defined as $L(\mathbf{w}) = \|\hat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\|^2$, where $\hat{\boldsymbol{\mu}}(\mathbf{w})$ represents the estimated mean vector parameterized by the weight vector $\mathbf{w}$, and $\boldsymbol{\mu}$ denotes the true mean vector. The associated risk, $R(\mathbf{w}) = \mathbb{E}[\|\hat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\|^2]$, quantifies the expected squared loss. An optimal weight matrix is identified by minimizing an approximation of $R(\mathbf{w})$, which serves as the model averaging criterion in this framework.

Assuming the covariance matrix $\boldsymbol{\Omega}$ is known, we propose a Mallows-type criterion for optimizing the weights $\mathbf{w}$ as follows:

$$C(\mathbf{w}) = \|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\|^2 + 2 \sum_{s=1}^{S} \sum_{h=1}^{H} \left\{ w_{s,h} \operatorname{tr}(\tilde{\mathbf{P}}_{s,h}\boldsymbol{\Omega}) + \frac{\partial \hat{\rho}_{s,h}}{\partial \mathbf{Y}^\top} \boldsymbol{\Omega} \frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\rho}_{s,h}} \mathbf{Y} + \frac{\partial \hat{\lambda}_{s,h}}{\partial \mathbf{Y}^\top} \boldsymbol{\Omega} \frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\lambda}_{s,h}} \mathbf{Y} \right\}. \tag{14}$$

This criterion balances model fit, captured by the first term $\|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\|^2$, against model complexity represented by the second term. To clarify the complexity term, we define the model degrees of freedom following Efron (2004) and Zou et al. (2007) as $df_{\mathbf{w}} = \operatorname{cov}(\hat{\boldsymbol{\mu}}(\mathbf{w})', \mathbf{Y}')$, where the covariance operator $\operatorname{cov}(\mathbf{a}, \mathbf{b}) = \mathbb{E}[(\mathbf{a} - \mathbb{E}[\mathbf{a}])(\mathbf{b} - \mathbb{E}[\mathbf{b}])']$ applies to arbitrary random vectors $\mathbf{a}$ and $\mathbf{b}$. The second term of $C(\mathbf{w})$ is shown to be an unbiased estimator of $df_{\mathbf{w}}$. Under the assumption of normally distributed errors ϵ, we apply Stein's lemma (Stein, 1981) to establish that:

$$\mathbb{E}[C(\mathbf{w})] = R(\mathbf{w}) - \operatorname{tr}(\boldsymbol{\Omega}), \tag{15}$$

demonstrating that $C(\mathbf{w})$ serves as an unbiased estimator of the risk, adjusted by a constant(see Appendix A).

In practice, the covariance matrix $\boldsymbol{\Omega}$ is typically unknown and requires estimation by $\hat{\boldsymbol{\Omega}}$. A feasible weight selection criterion is therefore proposed as follows:

$$\hat{C}_{\text{ori}}(\mathbf{w}) = \|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\|^2 + 2\sum_{s=1}^{S}\sum_{h=1}^{H} w_{s,h}\mathcal{D}_{s,h},$$

where the term "ori" refers to the original criterion. The complexity measure $\mathcal{D}_{s,h}$ is defined as:

$$\mathcal{D}_{s,h} = \text{tr}(\tilde{\mathbf{P}}_{s,h}\hat{\boldsymbol{\Omega}}) + \frac{\partial\hat{\rho}_{s,h}}{\partial\mathbf{Y}^\top}\hat{\boldsymbol{\Omega}}\frac{\partial\tilde{\mathbf{P}}_{s,h}}{\partial\hat{\rho}_{s,h}}\mathbf{Y} + \frac{\partial\hat{\lambda}_{s,h}}{\partial\mathbf{Y}^\top}\hat{\boldsymbol{\Omega}}\frac{\partial\tilde{\mathbf{P}}_{s,h}}{\partial\hat{\lambda}_{s,h}}\mathbf{Y}.$$

To ensure consistent selection of the true spatial weights matrix when it is a zero matrix, a penalty term is incorporated into $\hat{C}_{\text{ori}}(\mathbf{w})$, resulting in a modified criterion:

$$\hat{C}(\mathbf{w}) = \|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\|^2 + 2\sum_{s=1}^{S}\sum_{h=1}^{H} w_{s,h}\left\{\mathcal{D}_{s,h} + \frac{\text{pen}_{s,h}}{2}\right\}. \tag{16}$$

The penalty term $\text{pen}_{s,h}$ is defined as:

$$\text{pen}_{s,h} = \begin{cases} 0, & \text{if } s=1 \text{ and } h=1,\\ \frac{0.1}{\|\mathbf{X}\hat{\beta}_{1,h}-\mathbf{X}\hat{\beta}_{1,1}\|^2}, & \text{if } s=1, h\neq 1,\\ \frac{\sqrt{n(T-1)}}{\|\mathbf{X}\hat{\beta}_{s,1}-\mathbf{X}\hat{\beta}_{1,1}\|^2}, & \text{if } s\neq 1, h=1,\\ \frac{0.1}{\|\mathbf{X}\hat{\beta}_{1,h}-\mathbf{X}\hat{\beta}_{1,1}\|^2} + \frac{\sqrt{n(T-1)}}{\|\mathbf{X}\hat{\beta}_{s,1}-\mathbf{X}\hat{\beta}_{1,1}\|^2}, & \text{if } s\neq 1 \text{ and } h\neq 1. \end{cases}$$

This penalty term enhances the model selection criterion's ability to consistently identify the true spatial weights matrix by penalizing deviations from the baseline model ($s = 1, h = 1$). The rationale for this penalty structure relies on asymptotic behavior under specific regularity conditions. When the true spatial weights matrix $\mathbf{W}$ is the zero matrix, the estimators $\hat{\boldsymbol{\beta}}_{s,1}$ (for $s \neq 1$) and $\hat{\boldsymbol{\beta}}_{1,1}$ converge to the same limit, causing the term $\frac{\sqrt{n(T-1)}}{\|\mathbf{X}\hat{\beta}_{s,1}-\mathbf{X}\hat{\beta}_{1,1}\|^2}$ to approach infinity as $n(T-1) \to \infty$. Similarly, when the true matrix $\mathbf{M}$ is the zero matrix, $\hat{\boldsymbol{\beta}}_{1,h}$ (for $h \neq 1$) and $\hat{\boldsymbol{\beta}}_{1,1}$ share the same limit, leading to $\frac{0.1}{\|\mathbf{X}\hat{\beta}_{1,h}-\mathbf{X}\hat{\beta}_{1,1}\|^2} \to \infty$ as $n(T-1) \to \infty$. Conversely, when $\mathbf{W}$ is non-zero, $\hat{\boldsymbol{\beta}}_{s,1}$ (for $s \neq 1$) and $\hat{\boldsymbol{\beta}}_{1,1}$ generally have distinct limits, so the penalty term remains bounded, i.e., $\frac{\sqrt{n(T-1)}}{\|\mathbf{X}\hat{\beta}_{s,1}-\mathbf{X}\hat{\beta}_{1,1}\|^2} = O_p(1)$ as $n(T-1) \to \infty$. Likewise, when $\mathbf{M}$ is non-zero, $\hat{\boldsymbol{\beta}}_{1,h}$ (for $h \neq 1$) and $\hat{\boldsymbol{\beta}}_{1,1}$ have different limits, resulting in $\frac{0.1}{\|\mathbf{X}\hat{\beta}_{1,h}-\mathbf{X}\hat{\beta}_{1,1}\|^2} = O_p(1)$ as $n(T-1) \to \infty$. The selected weight is then

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{H}} \hat{C}(\mathbf{w}).$$

Define

$$\mathbf{H} = \begin{bmatrix} \tilde{\mathbf{P}}_{1,1}\mathbf{Y}-\mathbf{Y}, \tilde{\mathbf{P}}_{1,2}\mathbf{Y}-\mathbf{Y}, \dots, \tilde{\mathbf{P}}_{1,H}\mathbf{Y}-\mathbf{Y}\\ \tilde{\mathbf{P}}_{2,1}\mathbf{Y}-\mathbf{Y}, \tilde{\mathbf{P}}_{2,2}\mathbf{Y}-\mathbf{Y}, \dots, \tilde{\mathbf{P}}_{2,H}\mathbf{Y}-\mathbf{Y}\\ \tilde{\mathbf{P}}_{3,1}\mathbf{Y}-\mathbf{Y}, \tilde{\mathbf{P}}_{3,2}\mathbf{Y}-\mathbf{Y}, \dots, \tilde{\mathbf{P}}_{3,H}\mathbf{Y}-\mathbf{Y}\\ \dots,\\ \tilde{\mathbf{P}}_{S,1}\mathbf{Y}-\mathbf{Y}, \tilde{\mathbf{P}}_{S,2}\mathbf{Y}-\mathbf{Y}, \dots, \tilde{\mathbf{P}}_{S,H}\mathbf{Y}-\mathbf{Y} \end{bmatrix}.$$

and

$$\mathbf{L} = \begin{bmatrix} p\hat{\sigma}^2 & \text{trace}(\tilde{\mathbf{P}}_{1,2}\boldsymbol{\Omega}) & \dots & \text{trace}(\tilde{\mathbf{P}}_{1,H}\boldsymbol{\Omega})\\ \text{trace}(\tilde{\mathbf{P}}_{2,1}\boldsymbol{\Omega}) & \text{trace}(\tilde{\mathbf{P}}_{2,2}\hat{\Omega}) & \dots & \text{trace}(\tilde{\mathbf{P}}_{2,H}\hat{\Omega})\\ \vdots & \vdots & \ddots & \vdots\\ \text{trace}(\tilde{\mathbf{P}}_{S,1}\boldsymbol{\Omega}) & \text{trace}(\tilde{\mathbf{P}}_{S,2}\boldsymbol{\Omega}) & \dots & \text{trace}(\tilde{\mathbf{P}}_{S,H}\hat{\Omega}) \end{bmatrix} +$$

$$\begin{bmatrix} 0 & \frac{\partial\hat{\rho}_{1,1}}{\partial\mathbf{Y}^\top}\boldsymbol{\Omega}\frac{\partial\tilde{\mathbf{P}}_{1,1}}{\partial\hat{\rho}_{1,1}}\mathbf{Y} & \dots & \frac{\partial\hat{\rho}_{1,H}}{\partial\mathbf{Y}^\top}\boldsymbol{\Omega}\frac{\partial\tilde{\mathbf{P}}_{1,H}}{\partial\hat{\rho}_{1,H}}\mathbf{Y}\\ \frac{\partial\hat{\lambda}_{2,1}}{\partial\mathbf{Y}^\top}\boldsymbol{\Omega}\frac{\partial\tilde{\mathbf{P}}_{2,1}}{\partial\hat{\lambda}_{2,1}}\mathbf{Y} & \frac{\partial\hat{\rho}_{2,2}}{\partial\mathbf{Y}^\top}\hat{\Omega}\frac{\partial\tilde{\mathbf{P}}_{2,2}}{\partial\hat{\rho}_{2,2}}\mathbf{Y} + \frac{\partial\hat{\lambda}_{2,2}}{\partial\mathbf{Y}^\top}\hat{\Omega}\frac{\partial\tilde{\mathbf{P}}_{2,2}}{\partial\hat{\lambda}_{2,2}}\mathbf{Y} & \dots & \frac{\partial\hat{\rho}_{2,H}}{\partial\mathbf{Y}^\top}\hat{\Omega}\frac{\partial\tilde{\mathbf{P}}_{2,H}}{\partial\hat{\rho}_{2,H}}\mathbf{Y} + \frac{\partial\hat{\lambda}_{2,H}}{\partial\mathbf{Y}^\top}\hat{\Omega}\frac{\partial\tilde{\mathbf{P}}_{2,H}}{\partial\hat{\lambda}_{2,H}}\mathbf{Y}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial\hat{\lambda}_{S,1}}{\partial\mathbf{Y}^\top}\boldsymbol{\Omega}\frac{\partial\tilde{\mathbf{P}}_{S,1}}{\partial\hat{\lambda}_{S,1}}\mathbf{Y} & \frac{\partial\hat{\rho}_{S,2}}{\partial\mathbf{Y}^\top}\hat{\Omega}\frac{\partial\tilde{\mathbf{P}}_{S,2}}{\partial\hat{\rho}_{S,2}}\mathbf{Y} + \frac{\partial\hat{\lambda}_{S,2}}{\partial\mathbf{Y}^\top}\hat{\Omega}\frac{\partial\tilde{\mathbf{P}}_{S,2}}{\partial\hat{\lambda}_{S,2}}\mathbf{Y} & \dots & \frac{\partial\hat{\rho}_{S,H}}{\partial\mathbf{Y}^\top}\hat{\Omega}\frac{\partial\tilde{\mathbf{P}}_{S,H}}{\partial\hat{\rho}_{S,H}}\mathbf{Y} + \frac{\partial\hat{\lambda}_{S,H}}{\partial\mathbf{Y}^\top}\hat{\Omega}\frac{\partial\tilde{\mathbf{P}}_{S,H}}{\partial\hat{\lambda}_{S,H}}\mathbf{Y} \end{bmatrix} +$$

$$\begin{bmatrix} 0 & \frac{pen_{1,1}}{2} & \dots & \frac{pen_{1,H}}{2}\\ \frac{pen_{2,1}}{2} & \frac{pen_{2,2}}{2} & \dots & \frac{pen_{2,H}}{2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{pen_{S,1}}{2} & \frac{pen_{S,2}}{2} & \dots & \frac{pen_{S,H}}{2} \end{bmatrix}$$

It is straightforward to show that

$$\hat{C}(\mathbf{w}) = \mathbf{w}'\mathbf{H}'\mathbf{H}\mathbf{w} + 2\mathbf{w}'\mathbf{L}, \tag{17}$$

so that $\hat{C}(\mathbf{w})$ is a quadratic function of $\mathbf{w}$. Numerous software packages are available to obtain the solution to this problem (e.g., quadprog of Matlab), and they generally work effectively and efficiently, even when S, T are very large.

3.2. Asymptotic optimality

The model averaging estimator $\hat{\boldsymbol{\mu}}(\hat{\mathbf{w}})$ achieves asymptotic optimality, meaning its squared loss converges to that of the infeasible optimal averaging estimator. To formalize this, define the risk function:

$$R(\mathbf{w})^* = \mathbb{E}\left\| \sum_{s=1}^{S}\sum_{h=1}^{H} w_{s,h}\hat{\boldsymbol{\mu}}_{s,h}\Big|_{\rho_{s,h}=\rho^*_{s,h},\lambda_{s,h}=\lambda^*_{s,h}} - \boldsymbol{\mu}\right\|^2,$$

where $\hat{\boldsymbol{\mu}}_{s,h}$ is the estimated mean function evaluated at the limiting parameters $(\rho^*_{s,h}, \lambda^*_{s,h})$, and $\boldsymbol{\mu}$ is the true mean vector. Let $\xi_{n(T-1)} = \inf_{\mathbf{w}} R(\mathbf{w})^*$, representing the minimal achievable risk. Additionally, denote $\lambda_{\min}(\Pi)$ and $\lambda_{\max}(\Pi)$ as the smallest and largest singular values, respectively, of any square matrix Π. All asymptotic properties are established as $n(T-1) \to \infty$.

To ensure the asymptotic optimality of the model averaging estimator in the spatial panel data framework, we outline a comprehensive set of assumptions that govern risk bounds, error distributions, matrix properties, parameter smoothness, and model complexity. These assumptions are presented below, followed by a summary of their roles in achieving robust statistical inference as the sample size $n(T-1) \to \infty$.

Assumption 1. There exists a positive integer G such that the sum of inverse risks satisfies:

$$SH(\xi_{n(T-1)})^{-2G}\sum_{s=1}^{S}\sum_{h=1}^{H}(R^*_{s,h})^G = o(1).$$

Assumption 2. The error terms ε_i satisfy a finite moment condition of order at least $4G$, meaning there exists a constant $\kappa > 0$ such that for all components:

$$\mathbb{E}(|\varepsilon_i|^{4G}) \le \kappa < \infty.$$

Assumption 3. The true mean vector and covariance matrices are bounded as follows:

$$\|\boldsymbol{\mu}\|^2 = O(n(T-1)), \quad \lambda_{\max}(\boldsymbol{\Omega}) = O(1), \quad \text{and} \quad \lambda_{\max}(\hat{\boldsymbol{\Omega}}) = O_p(1).$$

Assumption 4. There exist constants $c_1 > 0$ and $c_2 > 0$ such that for all sample sizes:

$$c_1 \le \lambda_{\min}\left(\frac{\mathbf{X}'\mathbf{X}}{n(T-1)}\right) \le \lambda_{\max}\left(\frac{\mathbf{X}'\mathbf{X}}{n(T-1)}\right) \le c_2.$$

Assumption 5. Under the limiting parameters $(\rho^*_{s,h}, \lambda^*_{s,h})$, the spatial weight matrices satisfy:

$$\sup_{s,h}\lambda_{\max}\left((\mathbf{I}_{T-1}\otimes R_n(\rho^*_{s,h})^{-1})\right) = o(1); \quad \sup_{s,h}\lambda_{\max}\left((\mathbf{I}_{T-1}\otimes S_n(\lambda^*_{s,h})^{-1})\right) = o(1);$$

$$\sup_{s,h}\lambda_{\max}\left((\mathbf{I}_{T-1}\otimes \rho^*_{s,h}\mathbf{W}_s)\right) = o(1); \quad \sup_{s,h}\lambda_{\max}\left((\mathbf{I}_{T-1}\otimes \lambda^*_{s,h}\mathbf{M}_h)\right) = o(1).$$

Assumption 6. The parameter estimators exhibit smoothness, satisfying:

$$\xi^{-1}_{n(T-1)}\sup_{s,h}\left|\frac{\partial\hat{\rho}_{s,h}}{\partial\mathbf{Y}'}\hat{\boldsymbol{\Omega}}\frac{\partial\tilde{\mathbf{P}}_{s,h}}{\partial\hat{\rho}_{s,h}}\mathbf{Y} + \frac{\partial\hat{\lambda}_{s,h}}{\partial\mathbf{Y}'}\hat{\boldsymbol{\Omega}}\frac{\partial\tilde{\mathbf{P}}_{s,h}}{\partial\hat{\lambda}_{s,h}}\mathbf{Y}\right| = o(1).$$

Assumption 7. The projection operators converge as:

$$\frac{p}{\xi_{n(T-1)}} = o(1), \quad n(T-1)\xi^{-1}_{n(T-1)}\sup_{s,h}\lambda_{\max}(\tilde{\mathbf{P}}_{s,h} - \bar{\mathbf{P}}_{s,h}) = o(1).$$

Assumption 1 is a critical assumption, widely adopted in the model averaging literature (Liu and Okui, 2013; Wan et al., 2010). This component enables the determination of probability orders for terms that are linear and quadratic in $\tilde{\varepsilon}$ using the generalized Chebyshev's inequality and Whittle's inequality. It implicitly assumes a balance between the number of candidate models and the goodness-of-fit of individual models. In SARAR models, this assumption is crucial for maintaining estimation stability due to the spatial dependence in error terms. Assumption 2 guarantees that the error terms have finite higher-order moments, enabling the use of Markov-type inequalities to manage higher-order terms. This assumption is also common in the literature on the model selection optimality (Li, 1987; Zhang and Yu, 2018). It is used to show that some quadratic forms of the error terms required in the analysis are bounded in probability. This moment condition is applied to establish probabilistic bounds on both linear and quadratic terms of the error. In SARAR models, this assumption is crucial for maintaining estimation stability due to the spatial dependence in error terms. Assumption 3 ensures that the mean vector and covariance matrices remain bounded, maintaining stability in large samples. The sum of n elements of μ is also frequently employed in linear regression models (Liang et al., 2011). Extended from linear regression models, this assumption in spatial panels guarantees numerical stability after orthogonal transformation and prevents covariance explosion caused by spatial dependence. Assumption 4 guarantees that the covariate matrix $\mathbf{X}$ is well-conditioned, preventing degeneracy or explosion. In spatial panels, this is particularly important because spatial lag terms may introduce endogeneity, and well-conditioned covariates are prerequisites for QML estimation convergence. Assumption 5 ensures that the spatial weight matrices are invertible and bounded under limiting parameters, supporting numerical stability. More specifically, this assumption maintains the spectral radius of spatial transformation matrices below 1, which is crucial for model identifiability. Any violation of this condition would result in the divergence of QML estimates. Assumption 6 is a common assumption in the literature on model averaging (Liu and Okui, 2013). It is a high-level condition, ensures the smoothness of parameter estimators, reducing variability in estimation. This condition constrains the rate of change of spatial parameter derivatives, ensuring estimation consistency of complexity terms in Mallows' criterion and maintaining the unbiasedness of risk functions. Assumption 7 is a standard assumption in model averaging studies (Liu and Okui, 2013; Zhang and Yu, 2018). The first part of Assumption 7 stipulates that the parameter dimension grows more slowly than the minimum risk descent rate. The second component of Assumption 7 ensures that the term $n(T-1)\xi^{-1}_{n(T-1)} \sup \lambda_{\max}(\tilde{\mathbf{P}}_{s,h} - \mathbf{P}^*_{s,h}) = o_p(1)$, guaranteeing that the convergence rate of the projection matrix estimation error is strictly faster than the sample size growth, consistent with findings in Zhang et al. (2014). This assumption is central to the optimality of model averaging and manages the trade-off between model misspecification and asymptotic efficiency in spatial panels.

Theorem 3.1. *Under Assumptions 1–7, the model averaging estimator achieves asymptotic optimality, as expressed by*

$$\frac{L(\hat{\mathbf{w}})}{\inf_{\mathbf{w}\in\mathcal{H}} L(\mathbf{w})} \to 1 \tag{18}$$

in probability as $n(T-1) \to \infty$. The detailed proof of Theorem 1 is provided in the supplemental file. Furthermore, the proof implies that the estimator based on the original criterion is also asymptotically optimal. Specifically, with $\hat{\mathbf{w}}_{ori} = \arg\min_{\mathbf{w}\in\mathcal{H}} \hat{C}(\mathbf{w})_{ori}$, we have: $\frac{L(\hat{\mathbf{w}}_{ori})}{\inf_{\mathbf{w}\in\mathcal{H}} L(\mathbf{w})} \to 1$ in probability as $n(T-1) \to \infty$.

4. Analysis of degenerate models

This section examines the degenerate forms of the spatial panel SARAR model under specific conditions. By analyzing model specifications and estimation methods under different degenerate scenarios, we provide theoretical guidance for model selection in empirical research. We focus on three typical cases: (1) model with only spatial lag dependence; (2) model with only spatial error dependence; and (3) model without spatial effects. For each case, we systematically derive the simplified model form, parameter estimation methods, and corresponding model averaging criteria.

4.1. Model without spatial error dependence (SAR)

When the error term exhibits no spatial autocorrelation ($M = 0$) but the dependent variable retains spatial dependence, the model is specified as:

$$\mathbf{Y} = (\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))^{-1}\mathbf{V}, \tag{19}$$

where $\mathbf{S}_n(\lambda) = \mathbf{I}_n - \lambda\mathbf{W}$, and $\mathbf{W}$ is the spatial weight matrix. The covariance matrix simplifies to $\boldsymbol{\Omega} = \sigma_0^2(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))^{-1}(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda)')^{-1}$. The conditional expectation is $\boldsymbol{\mu} = (\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))^{-1}\mathbf{X}\boldsymbol{\beta}$.

Given the QML estimators (Lee and Yu, 2010) for the sth candidate model, the QML estimators of $\boldsymbol{\beta}$ and σ^2 are

$$\hat{\boldsymbol{\beta}}_s = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_s))\mathbf{Y},$$

and

$$\hat{\sigma}^2_s = \frac{1}{n(T-1)} \left\|(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_s))\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_s\right\|^2.$$

The estimator for the expected value $\boldsymbol{\mu}$ is $\hat{\boldsymbol{\mu}}_s = (\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_s))^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}_s = \tilde{\mathbf{P}}_s\mathbf{Y}$, where $\tilde{\mathbf{P}}_s = (\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_s))^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_s))$. For model averaging, let the weight vector $\mathbf{w} = (w_1, \ldots, w_S)' \in \mathcal{H} = \left\{\mathbf{w} \in [0,1]^S : \sum_{s=1}^S w_s = 1\right\}$. Define $\tilde{\mathbf{P}}(\mathbf{w}) = \sum_{s=1}^S w_s\tilde{\mathbf{P}}_s$, the model-averaged estimator is

$$\hat{\boldsymbol{\mu}}(\mathbf{w}) = \sum_{s=1}^S w_s\hat{\boldsymbol{\mu}}_s = \tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y}.$$

The model selection criterion, penalizing spatial lag complexity, is:

$$C(\mathbf{w})_1 = \left\|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\right\|^2 + 2\sum_{s=1}^S w_s \left\{\mathrm{tr}(\tilde{\mathbf{P}}_s\boldsymbol{\Omega}) + \frac{\partial\hat{\lambda}_s}{\partial\mathbf{Y}'}\boldsymbol{\Omega}\frac{\partial\tilde{\mathbf{P}}_s}{\partial\hat{\lambda}_s}\mathbf{Y}\right\}. \tag{20}$$

Defined squared loss as $L_1(\mathbf{w}) = \|\hat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\|^2$ and the associated risk is the expected squared loss given by $R_1(\mathbf{w}) = \mathbb{E}[\|\hat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\|^2]$. Under normality of ϵ and Stein's Lemma, the expected criterion satisfies:

$$\mathbb{E}[C(\mathbf{w})_1] = R(\mathbf{w}) - \mathrm{tr}(\boldsymbol{\Omega}),$$

In practice, $\boldsymbol{\Omega}$ is replaced by an estimator $\hat{\boldsymbol{\Omega}}$, yielding the feasible criterion:

$$\hat{C}(\mathbf{w})_1 = \left\|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\right\|^2 + 2\sum_{s=1}^{S} w_s \left\{ \mathrm{tr}(\tilde{\mathbf{P}}_s \hat{\boldsymbol{\Omega}}) + \frac{\partial \hat{\lambda}_s}{\partial \mathbf{Y}^\top} \hat{\boldsymbol{\Omega}} \frac{\partial \tilde{\mathbf{P}}_s}{\partial \hat{\lambda}_s} \mathbf{Y} + \mathrm{pen}_s/2 \right\}. \tag{21}$$

The penalty term for model selection is defined as:

$$\mathrm{pen}_s = \begin{cases} \sqrt{n(T-1)}/\left\|\mathbf{X}\hat{\boldsymbol{\beta}}_s - \mathbf{X}\hat{\boldsymbol{\beta}}_1\right\|^2, & s \geq 2, \\ 0, & s = 1, \end{cases}$$

where $\mathbf{X}$ is the design matrix, $\hat{\boldsymbol{\beta}}_s$ is the estimated coefficient vector for the sth model, and $\hat{\boldsymbol{\beta}}_1$ corresponds to the baseline model ($s = 1$). The rationale for this penalty term is as follows: when the true coefficient matrix is the zero matrix, the estimators $\hat{\boldsymbol{\beta}}_s$ for $s \geq 2$ and $\hat{\boldsymbol{\beta}}_1$ should have the same asymptotic limits, leading to $\mathrm{pen}_s \to \infty$ as $n(T-1) \to \infty$. Conversely, when the true coefficient matrix is not the zero matrix, $\hat{\boldsymbol{\beta}}_s$ for $s \geq 2$ and $\hat{\boldsymbol{\beta}}_1$ have different limits, resulting in $\mathrm{pen}_s = O_p(1)$ as $n(T-1) \to \infty$.

Theorem 4.1. *Under Assumptions 1–7, the selected weights $\hat{\mathbf{w}}$ achieve asymptotic optimality:*

$$\frac{L_1(\hat{\mathbf{w}})}{\inf_{\mathbf{w}} L_1(\mathbf{w})} \xrightarrow{p} 1 \quad \text{as } n(T-1) \to \infty, \tag{22}$$

ensuring minimal risk in large samples. The proof process is similar to that of Theorem 3.1.

4.2. Model without spatial lag dependence (SEM)

When the dependent variable exhibits no spatial dependence ($\mathbf{W} = 0$) but the error term shows spatial autocorrelation, the model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho))^{-1}\mathbf{V}, \tag{23}$$

where $\mathbf{R}_n(\rho) = \mathbf{I}_n - \rho\mathbf{M}$, and $\mathbf{M}$ is the error spatial weight matrix. The covariance matrix is $\boldsymbol{\Omega} = \sigma_0^2(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho))^{-1}(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho)')^{-1}$. and the conditional expectation is $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$.

Given the QML estimators (Lee and Yu, 2010) for the hth candidate model, the estimators of $\boldsymbol{\beta}$ and σ_0^2 are

$$\hat{\boldsymbol{\beta}}_h = \left(\mathbf{X}'(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_h))'(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_h))\mathbf{X}\right)^{-1} \mathbf{X}'(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_h))\mathbf{Y},$$

and $\hat{\sigma}_h^2 = \frac{1}{n(T-1)} \left\|(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_h))\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_h\right\|^2$. The estimator for $\boldsymbol{\mu}$ is $\hat{\boldsymbol{\mu}}_h = \mathbf{X}\hat{\boldsymbol{\beta}}_h = \tilde{\mathbf{P}}_h\mathbf{Y}$, where $\tilde{\mathbf{P}}_h = \mathbf{X}\left(\mathbf{X}'(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_h))'(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_h))\mathbf{X}\right)^{-1} \mathbf{X}'(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_h))$.

For model averaging, let $\mathbf{w} = (w_1, \ldots, w_H)' \in \mathcal{H} = \left\{\mathbf{w} \in [0,1]^H : \sum_{h=1}^H w_h = 1\right\}$. Define $\tilde{\mathbf{P}}(\mathbf{w}) = \sum_{h=1}^H w_h \tilde{\mathbf{P}}_h$, the model-averaged estimator is

$$\hat{\boldsymbol{\mu}}(\mathbf{w}) = \sum_{h=1}^{H} w_h \hat{\boldsymbol{\mu}}_h = \tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y}.$$

The model selection criterion is

$$C(\mathbf{w})_2 = \left\|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\right\|^2 + 2\sum_{h=1}^{H} w_h \left\{ \mathrm{tr}(\tilde{\mathbf{P}}_h \boldsymbol{\Omega}) + \frac{\partial \hat{\rho}_h}{\partial \mathbf{Y}'} \boldsymbol{\Omega} \frac{\partial \tilde{\mathbf{P}}_h}{\partial \hat{\rho}_h} \mathbf{Y} \right\}.$$

Defined squared loss as $L_2(\mathbf{w}) = \|\hat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\|^2$ and the associated risk is the expected squared loss, given by $R_2(\mathbf{w}) = \mathbb{E}[\|\hat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\|^2]$. Under normality and Stein's Lemma, the expected criterion is:

$$\mathbb{E}[C(\mathbf{w})_2] = R(\mathbf{w}) - \mathrm{tr}(\boldsymbol{\Omega}).$$

With $\boldsymbol{\Omega}$ estimated by $\hat{\boldsymbol{\Omega}}$, the feasible criterion is

$$\hat{C}(\mathbf{w})_2 = \left\|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\right\|^2 + 2\sum_{h=1}^{H} w_h \left\{ \mathrm{tr}(\tilde{\mathbf{P}}_h \hat{\boldsymbol{\Omega}}) + \frac{\partial \hat{\rho}_h}{\partial \mathbf{Y}'} \hat{\boldsymbol{\Omega}} \frac{\partial \tilde{\mathbf{P}}_h}{\partial \hat{\rho}_h} \mathbf{Y} + \mathrm{pen}_s/2 \right\}. \tag{24}$$

The penalty term for model selection is defined as:

$$\mathrm{pen}_h = \begin{cases} 0.1/\left\|\mathbf{X}\hat{\boldsymbol{\beta}}_h - \mathbf{X}\hat{\boldsymbol{\beta}}_1\right\|^2, & h \geq 2, \\ 0, & h = 1, \end{cases}$$

where $\mathbf{X}$ is the design matrix, $\hat{\boldsymbol{\beta}}_h$ is the estimated coefficient vector for the hth model, and $\hat{\boldsymbol{\beta}}_1$ corresponds to the baseline model ($h = 1$). The rationale for this penalty term is as follows: when the true coefficient matrix is the zero matrix, the estimators $\hat{\boldsymbol{\beta}}_h$ for $h \geq 2$ and $\hat{\boldsymbol{\beta}}_1$ should have the same asymptotic limits, leading to $\text{pen}_h \to \infty$ as $n(T-1) \to \infty$. Conversely, when the true coefficient matrix is not the zero matrix, $\hat{\boldsymbol{\beta}}_h$ for $h \geq 2$ and $\hat{\boldsymbol{\beta}}_1$ have different limits, resulting in $\text{pen}_h = O_p(1)$ as $n(T-1) \to \infty$.

Theorem 4.2. *Under Assumptions 1–7, the selected weights $\hat{\mathbf{w}}$ achieve asymptotic optimality:*

$$\frac{L_2(\hat{\mathbf{w}})}{\inf_{\mathbf{w}} L_2(\mathbf{w})} \xrightarrow{p} 1 \quad \text{as } n(T-1) \to \infty, \tag{25}$$

guaranteeing optimal risk performance in large samples. The proof process is similar to that of Theorem 3.1.

4.3. Model without spatial effects ($\mathbf{W} = 0, \mathbf{M} = 0$)

When neither the dependent variable nor the error term exhibits spatial dependence ($\mathbf{W} = 0, \mathbf{M} = 0$), the model reduces to the classical panel data model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{V}. \tag{26}$$

The maximum likelihood estimator (MLE) for $\boldsymbol{\beta}$ is: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, and the error variance estimator is: $\hat{\sigma_0}^2 = \frac{1}{n(T-1)} \left\| \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \right\|^2$. The projection matrix is: $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and the conditional expectation is $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} = \mathbf{P}\mathbf{Y}$.

The model averaging criterion simplifies to:

$$\hat{C}_{\text{ori},1} = \|\mathbf{P}\mathbf{Y} - \mathbf{Y}\|^2 + 2p\hat{\sigma}^2, \tag{27}$$

which aligns with Mallows' C_p criterion. This confirms the method's appropriateness in the absence of spatial effects. When the candidate model set includes the non-spatial model, the penalty mechanism favors this simpler model, avoiding unnecessary complexity.

By incorporating mean structures, projection operators, and variance–covariance estimation into the model averaging framework, this approach effectively addresses both spatial and non-spatial specifications. The adaptive penalty mechanism maintains consistency with classical methods in the absence of spatial effects while adeptly capturing spatial dependence when it exists. This dual capability strengthens theoretical rigor and enhances empirical applicability in spatial econometrics.

5. Monte Carlo experiments

To rigorously evaluate the performance of the proposed model averaging (MA) framework for spatial panel SARAR models, we employ Monte Carlo simulations to assess its effectiveness in addressing uncertainty in spatial weight matrices, achieving accurate parameter estimation, and ensuring robustness across a range of spatial dependence structures. This section is organized as follows: Section 5.1 investigates the full SARAR model across diverse scenarios; Section 5.2 analyzes the performance of degenerate SAR and SEM specifications; and Section 5.3 compares MA with classical model selection criteria, including Mallows' C_p, Akaike information criterion (AIC), and Bayesian information criterion (BIC); Section 5.4 extends the analysis to irregular spatial weight matrices; and Section 5.5 evaluates computational efficiency.

5.1. Simulation analysis of the SARAR model

5.1.1. Goals and design

The Monte Carlo experiments are designed to evaluate the performance of the proposed model averaging (MA) framework for spatial autoregressive panel data models (SARAR) under varying spatial dependence structures. The objectives are to: (a) verify the MA method's capability to address uncertainty in multiple spatial weight matrices; (b) evaluate the estimation accuracy of parameters λ, ρ, and $\boldsymbol{\beta}$; (c) analyze robustness across no-effect, single-effect, and mixed-effect spatial dependence structures; (d) confirm the asymptotic optimality of squared loss; and (e) test the adaptive penalty mechanism's efficiency in identifying the true model; and (f) assess model fitting diagnostics including residual distributions, actual-fitted correspondence.

The data-generating process (DGP) adheres to the SARAR model specified in Section 2:

$$\mathbf{Y}_{nt} = \lambda \mathbf{W}_n \mathbf{Y}_{nt} + \mathbf{X}_{nt}\boldsymbol{\beta} + \mathbf{c}_{n0} + \mathbf{U}_{nt}, \quad \mathbf{U}_{nt} = \rho \mathbf{M}_n \mathbf{U}_{nt} + \mathbf{V}_{nt}, \quad t = 1, \ldots, T,$$

where $\mathbf{V}_{nt} \sim \text{i.i.d.}\mathcal{N}(0, \sigma_0^2 \mathbf{I}_n)$, $\mathbf{X}_{nt} \sim \mathcal{N}(0, 1)$ with $k = 3$ covariates, and true coefficients $\boldsymbol{\beta} = [1.5, -2.0, 0.5]'$. Individual fixed effects $\mathbf{c}_{n0}$ are eliminated via the orthogonal transformation $\mathbf{Q} = \mathbf{F}'_{T,T-1} \otimes \mathbf{I}_n$. The response variable is generated via the reduced form:

$$\mathbf{Y} = \left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda)\right)^{-1} \left(\mathbf{X}\boldsymbol{\beta} + \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho)\right)^{-1} \mathbf{V} \right),$$

where $\mathbf{S}_n(\lambda) = \mathbf{I}_n - \lambda \mathbf{W}_n$ and $\mathbf{R}_n(\rho) = \mathbf{I}_n - \rho \mathbf{M}_n$. The experimental design includes cross-sectional units $n \in \{100, 225\}$ on a square grid, time dimensions $T \in \{5, 10\}$. Each configuration is replicated $N_{\text{sim}} = 1000$ times using MATLAB, with spatial weight matrices generated from grid coordinates, QML estimators optimized via `fminunc`, and numerical stability ensured. The QML estimation is initialized via 2SLS for λ and GMM for ρ. The detailed code is provided in the supplemental file.

Table 1
Model averaging performance for SARAR in Setting 1: True matrices in candidate set.

T = 5	n	Loss	MA weight	RMSE of $\hat{\lambda}$	RMSE of $\hat{\rho}$	RMSE of $\hat{\beta}$
No spatial effects	100	0.3036	1	0.0000	0.9000	0.0137
(W1,M1)	225	0.4429	1	0.0000	0.9000	0.0079
SAR Model	100	1.9083	0.8857	0.0181	0.8902	0.0256
(W2,M1)	225	2.5031	0.9744	0.0124	0.8011	0.0179
SEM Model	100	1.9923	0.5310	0.2335	0.0606	0.0328
(W1,M4)	225	2.4745	0.516	0.2346	0.0532	0.0228
Full SARAR Model	100	10.6820	0.674	0.0294	0.2316	0.0289
(W3,M2)	225	15.5653	0.7473	0.0238	0.1706	0.0206
T = 10	n	Loss	MA weight	RMSE of $\hat{\lambda}$	RMSE of $\hat{\rho}$	RMSE of $\hat{\beta}$
No spatial effects	100	0.2730	1	0.0000	0.9000	0.0089
(W1,M1)	225	0.3668	1	0.0000	0.9000	0.0051
SAR Model	100	2.3915	0.9335	0.0062	0.8855	0.0148
(W2,M1)	225	2.3731	0.9832	0.0148	0.8476	0.0166
SEM Model	100	1.5022	0.5777	0.2096	0.0492	0.0198
(W1,M4)	225	1.6603	0.6130	0.2477	0.0513	0.0142
Full SARAR Model	100	13.1592	0.7187	0.0293	0.1714	0.0197
(W3,M2)	225	9.2017	0.8342	0.0146	0.1033	0.0165

Performance is quantified by: (1) prediction accuracy via squared loss $\|\hat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\|^2$, (2) parameter estimation accuracy via root mean squared error (RMSE) of $\hat{\lambda}$, $\hat{\rho}$, and $\hat{\boldsymbol{\beta}}$, and (3) model selection accuracy via MA weights $\hat{w}_{s,h}$. Additionally, model diagnostics are assessed through "Actual vs. Fitted Values" plots and "Histogram of Residuals" to evaluate the model's fit and residual distribution.

Two distinct experimental settings are considered to evaluate the MA framework's robustness under different conditions for the true and candidate spatial weight matrices.

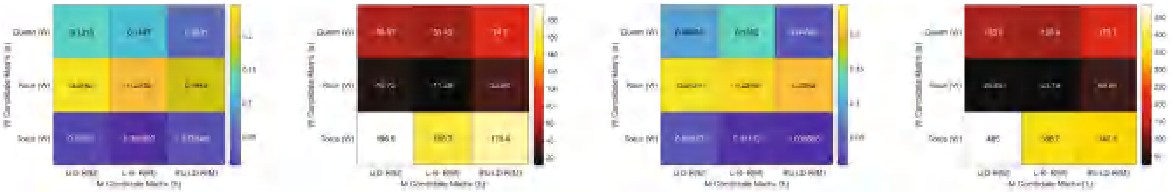
Setting 1: True matrices in candidate set with zero matrices. In the first setting, the true spatial weight matrices are included in the candidate set, which also contains zero matrices to allow for no spatial effects. The candidate spatial weight matrices include four $\mathbf{W}$ matrices: zero matrix ($\mathbf{W}_1$), Queen contiguity ($\mathbf{W}_2$), Rook contiguity ($\mathbf{W}_3$), and torus matrix ($\mathbf{W}_4$), and four $\mathbf{M}$ matrices: zero matrix ($\mathbf{M}_1$), vertical Rook ($\mathbf{M}_2$), horizontal Rook ($\mathbf{M}_3$), and diagonal Rook ($\mathbf{M}_4$), all row-standardized, yielding $4 \times 4 = 16$ model combinations. Four scenarios are considered: (1). No spatial effects ($\lambda = 0, \rho = 0, \mathbf{W}_1, \mathbf{M}_1$), (2). SAR model ($\lambda = 0.5, \rho = 0, \mathbf{W}_2, \mathbf{M}_1$), (3). SEM model ($\lambda = 0, \rho = 0.7, \mathbf{W}_1, \mathbf{M}_4$), (4). Full SARAR model ($\lambda = 0.5, \rho = 0.7, \mathbf{W}_3, \mathbf{M}_2$).

Setting 2: True matrices not in candidate set without zero matrices. In the second setting, the true spatial weight matrices are not included in the candidate set, and the candidate set excludes zero matrices to test the MA method's ability to approximate complex spatial dependencies. The true matrices are defined as mixed structures: $\mathbf{W}_{\text{true}} = \mathbf{W}_2 + \mathbf{W}_3$ for the dependent variable and $\mathbf{M}_{\text{true}} = \mathbf{M}_2 + \mathbf{M}_3$ for the error term, both row-standardized. The candidate set is restricted to non-zero matrices: $\mathbf{W}_2$ (Queen contiguity), $\mathbf{W}_3$ (Rook contiguity), $\mathbf{W}_4$ (torus matrix) for $\mathbf{W}$, and $\mathbf{M}_2$ (vertical Rook), $\mathbf{M}_3$ (horizontal Rook), $\mathbf{M}_4$ (diagonal Rook) for $\mathbf{M}$, yielding $3 \times 3 = 9$ model combinations. The DGP parameters remain as in Setting 1, with $\lambda = 0.5$, $\rho = 0.7$, and $\sigma_0^2 = 0.5$. The experiments are conducted for $n \in \{100, 225\}$ and $T \in \{5, 10\}$, with 1000 replications.

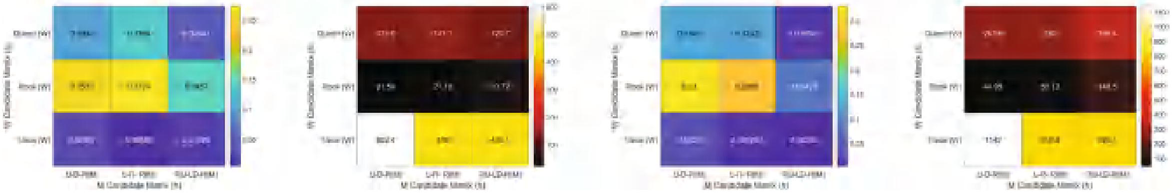
5.1.2. *Experimental results*

Table 1 presents the performance of the MA approach across the four scenarios. In the no spatial effects scenario ($\mathbf{W}_1, \mathbf{M}_1$), the MA method assigns full weight (1.0000) to the correct non-spatial model, achieving zero RMSE for $\hat{\lambda}$, low RMSE for $\hat{\boldsymbol{\beta}}$ (e.g., 0.0051 at $T = 10, n = 225$), and minimal loss (0.2730–0.4429), highlighting the adaptive penalty's effectiveness in detecting the absence of spatial effects. For the SAR model ($\mathbf{W}_2, \mathbf{M}_1$), MA strongly favors the correct spatial lag matrix $\mathbf{W}_2$ with weights of 0.8857–0.9832, reducing the RMSE of $\hat{\lambda}$ by approximately 40% from $T = 5$ to $T = 10$ (e.g., from 0.0181 to 0.0062 at $n = 100$), with moderate loss (1.9083–2.5031) reflecting spatial lag complexity. In the SEM model ($\mathbf{W}_1, \mathbf{M}_4$), MA assigns lower weights to the correct error matrix $\mathbf{M}_4$ (0.5160–0.6130), indicating challenges in modeling spatial error structures, yet maintains low RMSE for $\hat{\rho}$ ($\leq$0.0606) despite high RMSE for the irrelevant $\hat{\lambda}$, with loss ranging from 1.5022 to 2.4745. The full SARAR model ($\mathbf{W}_3, \mathbf{M}_2$), with dual spatial dependence, exhibits the highest loss (9.2017–15.5653); however, increasing T to 10 reduces the RMSE of $\hat{\rho}$ by 55% (from 0.2316 to 0.1033 at $n = 225$), and MA weights reach 0.8342 for the correct model, showcasing robustness. Key insights include: (1) extending the time dimension T significantly enhances estimation precision, particularly for SEM and SARAR models; (2) the adaptive penalty mechanism reliably identifies non-spatial models; and (3) estimating complex spatial error structures remains challenging, with $\hat{\rho}$ RMSE often exceeding 0.1. These results affirm the MA framework's robustness and adaptability across diverse spatial scenarios.

Fig. 1 and Table 2, report results for the mixed spatial structure scenario ($\mathbf{W} = \mathbf{W}_2 + \mathbf{W}_3$, $\mathbf{M} = \mathbf{M}_2 + \mathbf{M}_3$). At $T = 10, n = 225$, over 80% of MA weights concentrate on the true matrix combinations (Fig. 1(g)), with MA loss (10.0042) significantly lower than individual candidate losses, demonstrating robustness to complex dependencies. Parameter estimation shows high accuracy for $\hat{\lambda}$ (RMSE 0.0180) and $\hat{\boldsymbol{\beta}}$ (RMSE $\leq$ 0.0166), but $\hat{\rho}$ exhibits higher RMSE (0.3097–0.3829) due to the intricate error structure. Increasing T from 5 to 10 enhances weight concentration and reduces loss, though larger n at $T = 5$ slightly increases loss (15.0031) due



(a) $T = 5, n = 100$, Weights (b) $T = 5, n = 100$, Loss (c) $T = 10, n = 100$, Weights (d) $T = 10, n = 100$, Loss

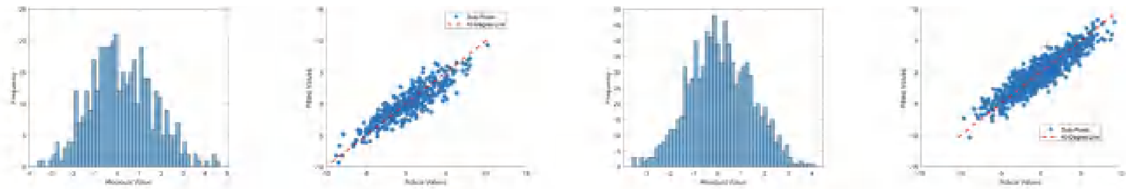


(e) $T = 5, n = 225$, Weights (f) $T = 5, n = 225$, Loss (g) $T = 10, n = 225$, Weights (h) $T = 10, n = 225$, Loss

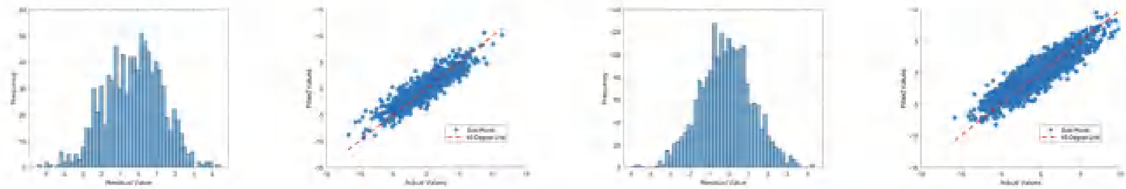
Fig. 1. Model averaging performance for SARAR in Setting 2: True matrices $\mathbf{W}_2 + \mathbf{W}_3, \mathbf{M}_2 + \mathbf{M}_3$.

Table 2
Model averaging performance for SARAR in Setting 2: True matrices $\mathbf{W}_2 + \mathbf{W}_3, \mathbf{M}_2 + \mathbf{M}_3$.

Scenario	Avg. Loss	RMSE of $\hat{\lambda}$	RMSE of $\hat{\rho}$	RMSE of $\hat{\beta}$
$T = 5, n = 100$	8.8056	0.0309	0.3529	0.0298
$T = 5, n = 225$	15.0031	0.0309	0.3829	0.0205
$T = 10, n = 100$	13.8559	0.0380	0.3206	0.0222
$T = 10, n = 225$	10.0042	0.0180	0.3097	0.0166



(a) $T = 5, n = 100$, Histogram of Residuals (b) $T = 5, n = 100$, Actual vs. Fitted Values (c) $T = 10, n = 100$, Histogram of Residuals (d) $T = 10, n = 100$, Actual vs. Fitted Values



(e) $T = 5, n = 225$, Histogram of Residuals (f) $T = 5, n = 225$, Actual vs. Fitted Values (g) $T = 10, n = 225$, Histogram of Residuals (h) $T = 10, n = 225$, Actual vs. Fitted Values

Fig. 2. Histogram of residuals and Actual vs. Fitted values SARAR in Setting 2: True matrices $\mathbf{W}_2 + \mathbf{W}_3, \mathbf{M}_2 + \mathbf{M}_3$.

to heightened complexity. These findings underscore the MA method's efficacy in handling composite spatial dependencies, with potential for further refinement in error structure estimation.

To further address model diagnostics, Fig. 2 present the histograms of residuals and scatter plots of actual versus fitted values for the mixed spatial structure scenario. The residuals exhibit approximate normality, centered around zero with symmetric distributions, indicating good model fit. The actual versus fitted values align closely along the 45-degree line, with minor deviations at extreme values, confirming the MA framework's predictive accuracy. These visuals substantiate the simulation results, showing reduced variance in residuals as n and T increase, aligning with asymptotic optimality.

Table 3
Model averaging performance for SAR: True matrices in candidate set.

T = 5		True W is $\mathbf{W}_1$				True W is $\mathbf{W}_2$			
		$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$
n = 100	RMSE of $\hat{\lambda}$	0.0000	0.0091	0.0052	0.0021	0.0000	0.0022	0.1935	0.2996
	RMSE of $\hat{\beta}$	0.0065	0.0064	0.0067	0.0067	0.0598	0.0053	0.0343	0.0284
	MA weights	1.0000	0	0	0	0.0926	0.9076	0.0000	0.0000
	Loss	0.0590	0.0953	0.0804	0.0622	172.7	6.369	84.1	139.8
	MA loss	0.0368				11.7289			
n = 225	RMSE of $\hat{\lambda}$	0.0000	0.0007	0.0012	0.0015	0.0000	0.0033	0.1970	0.2870
	RMSE of $\hat{\beta}$	0.0021	0.0021	0.0021	0.0019	0.0525	0.0028	0.0259	0.0318
	MA weights	1.0000	0	0	0	0.0716	0.9282	0.0002	0.0000
	Loss	0.0139	0.0145	0.0170	0.0246	423.7	11.21	169.5	314.3
	MA loss	0.0452				14.4523			
T = 10		True W is $\mathbf{W}_1$				True W is $\mathbf{W}_2$			
		$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$
n = 100	RMSE of $\hat{\lambda}$	0.0000	0.0055	0.0008	0.0013	0.0000	0.0027	0.2159	0.2841
	RMSE of $\hat{\beta}$	0.0032	0.0033	0.0032	0.0033	0.0536	0.0034	0.0304	0.0339
	MA weights	1.0000	0	0	0	0.0674	0.9325	0.0000	0.0000
	Loss	0.0315	0.0571	0.0326	0.0351	347.6	8.579	162.0	251.7
	MA loss	0.0320				11.7289			
n = 225	RMSE of $\hat{\lambda}$	0.0000	0.0003	0.0002	0.0008	0.0000	0.0033	0.1956	0.2972
	RMSE of $\hat{\beta}$	0.0015	0.0015	0.0015	0.0016	0.0482	0.0017	0.0179	0.0252
	MA weights	1.0000	0	0	0	0.0657	0.9341	0.0002	0.0000
	Loss	0.0171	0.0172	0.0173	0.0219	823.3	16.26	332.0	641.7
	MA loss	0.0328				25.8517			

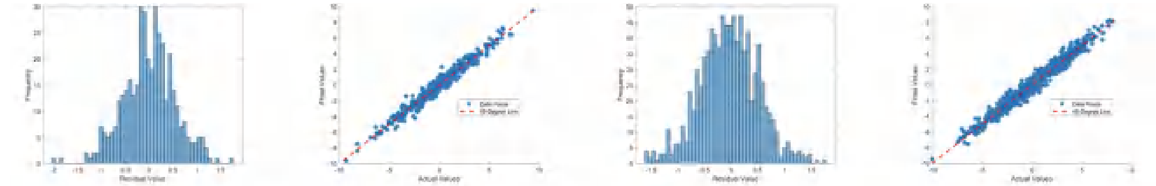
Table 4
Model averaging performance for SAR: True matrix $W2 + W3$.

	T = 5	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	MA
$n = 100$	RMSE of $\hat{\lambda}$	0.0000	0.0027	0.0991	0.2668	0.0473
	RMSE of $\hat{\beta}$	0.0769	0.0322	0.0113	0.0204	0.0069
	Loss	196.8	27.2	33.1	151.3	13.3089
	MA weights	0.0084	0.4746	0.5168	0.0002	
$n = 225$	RMSE of $\hat{\lambda}$	0.0000	0.0129	0.0914	0.2399	0.0404
	RMSE of $\hat{\beta}$	0.0660	0.0154	0.0136	0.0256	0.0085
	Loss	522	65.87	58.75	334.5	27.1051
	MA weights	0.025	0.4979	0.4992	0.0004	
	T = 10	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	MA
$n = 100$	RMSE of $\hat{\lambda}$	0.0000	0.0021	0.0941	0.24667	0.0436
	RMSE of $\hat{\beta}$	0.0625	0.0209	0.0073	0.0405	0.0092
	Loss	536.3	77.91	85.35	368.2	25.0403
	MA weights	0.0014	0.4948	0.5038	0.0000	
$n = 225$	RMSE of $\hat{\lambda}$	0.0000	0.0019	0.0983	0.2570	0.0416
	RMSE of $\hat{\beta}$	0.0625	0.0209	0.0073	0.0405	0.0075
	Loss	1037	134.6	144.6	683.6	47.5184
	MA weights	0.0001	0.4971	0.5028	0.000	

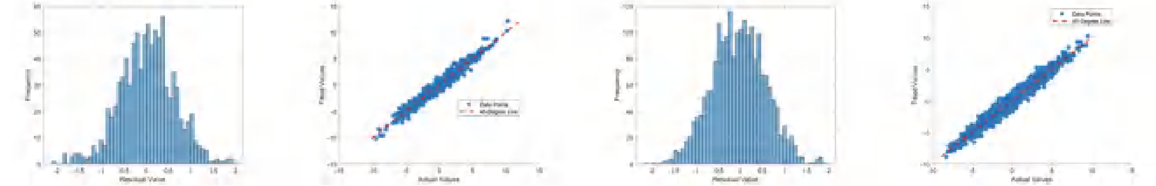
5.2. Degenerate model analysis

5.2.1. SAR model

The MA approach is evaluated for the SAR model, a degenerate form of SARAR with $\rho = 0$, employing the experimental design from Section 5.1.1 and candidate matrices $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{W}_4$. Scenarios include no spatial effects ($\lambda = 0, \mathbf{W}_1$), single spatial structure ($\lambda = 0.5, \mathbf{W}_2$), and mixed structure ($\mathbf{W}_2 + \mathbf{W}_3$). Table 3 shows results for single-structure scenarios: for $\mathbf{W}_1$, MA assigns full weight (1.0000), with zero RMSE for $\hat{\lambda}$ and low RMSE for $\hat{\beta}$ (e.g., 0.0015 at $T = 10, n = 225$), and minimal MA loss (0.0320–0.0452); for $\mathbf{W}_2$, MA weights reach 0.9341 ($T = 10, n = 225$), with low RMSE ($\hat{\lambda}$: 0.0033, $\hat{\beta}$: 0.0017) and MA loss (11.7289–25.8517) significantly below misspecified models. Table 4 addresses the mixed structure, where MA weights are nearly equal for $\mathbf{W}_2$ and $\mathbf{W}_3$ (e.g., 0.4971, 0.5028 at $T = 10, n = 225$), excluding irrelevant matrices, with MA loss (e.g., 47.5184) lower than individual losses (e.g., 1037 for $\mathbf{W}_1$), highlighting the method's adaptability to complex spatial dependencies. Fig. 3 for the mixed spatial structure scenario ($\mathbf{W}_2 + \mathbf{W}_3$) demonstrate that the residuals are approximately normally distributed, centered at zero, and symmetric, indicating a good model fit. The scatter plots of actual versus fitted values show a strong alignment along the 45-degree line, with minor deviations at extreme values, confirming the predictive accuracy of the MA framework.



(a) $T = 5, n = 100$, Histogram of Residuals (b) $T = 5, n = 100$, Actual vs. Fitted Values (c) $T = 10, n = 100$, Histogram of Residuals (d) $T = 10, n = 100$, Actual vs. Fitted Values



(e) $T = 5, n = 225$, Histogram of Residuals (f) $T = 5, n = 225$, Actual vs. Fitted Values (g) $T = 10, n = 225$, Histogram of Residuals (h) $T = 10, n = 225$, Actual vs. Fitted Values

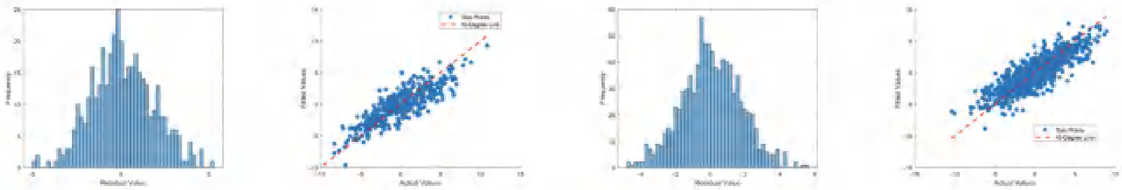
Fig. 3. Histogram of residuals and Actual vs. Fitted values for SAR: True matrices $\mathbf{W}_2 + \mathbf{W}_3$.

Table 5
Model averaging performance for SEM.

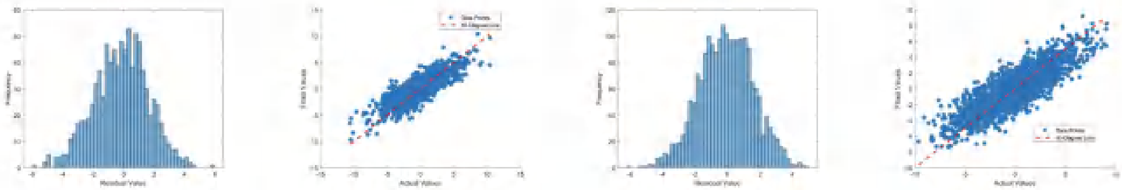
T = 5		True **M** is $\mathbf{M}_3$			True **M** is $\mathbf{M}_2 + \mathbf{M}_3$		
		$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$	$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$
n = 100	RMSE of $\hat{\rho}$	0.9348	0.0015	0.9114	0.2255	0.2245	0.3957
	RMSE of $\hat{\boldsymbol{\beta}}$	0.2725	0.0561	0.2796	0.0658	0.0663	0.0688
	MA weights	0.0167	0.9643	0.0191	0.3809	0.4091	0.2091
	Loss	119.9737	5.1553	133.9949	8.4881	7.4038	8.2330
	MA loss		19.1765			7.1814	
n = 225	RMSE of $\hat{\rho}$	0.8420	0.0031	0.8551	0.2258	0.2263	0.4021
	RMSE of $\hat{\boldsymbol{\beta}}$	0.3105	0.0310	0.3065	0.0466	0.0455	0.0501
	MA weights	0.0382	0.8775	0.0844	0.4941	0.2855	0.2203
	Loss	341.7	9.343	308.8	10.2421	7.9955	9.6258
	MA loss		45.4741			8.9202	
T = 10		True **M** is $\mathbf{M}_3$			True **M** is $\mathbf{M}_2 + \mathbf{M}_3$		
		$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$	$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$
n = 100	RMSE of $\hat{\rho}$	0.9296	0.0046	0.9331	0.2196	0.2159	0.3905
	RMSE of $\hat{\boldsymbol{\beta}}$	0.1004	0.0342	0.1013	0.0429	0.0449	0.0463
	MA weights	0.8918	0.8688	0.0420	0.4374	0.3659	0.1967
	Loss	31.8830	3.6366	30.8502	7.8075	6.8633	7.5255
	MA loss		56.6645			6.5630	
n = 225	RMSE of $\hat{\rho}$	0.9594	0.0020	0.9602	0.2294	0.2229	0.3990
	RMSE of $\hat{\boldsymbol{\beta}}$	0.2666	0.0140	0.2663	0.0335	0.0326	0.0362
	MA weights	0.0847	0.8169	0.0985	0.4666	0.4315	0.1019
	Loss	456.4304	1.3029	458.3563	9.2651	8.2181	9.7469
	MA loss		71.0281			7.8890	

5.2.2. SEM model

For the SEM model ($\lambda = 0, \rho = 0.5$), we use candidate matrices $\mathbf{M}_2, \mathbf{M}_2, \mathbf{M}_4$, testing single ($\mathbf{M}_3$) and mixed ($\mathbf{M}_2 + \mathbf{M}_3$) structures. Table 5 shows that for $\mathbf{M}_3$, MA assigns dominant weights (e.g., 0.9643 at $T = 5, n = 100$), with low RMSE for $\hat{\rho}$ (0.0020) and $\hat{\boldsymbol{\beta}}$ (0.0140), and MA loss (e.g., 71.0281) controlled compared to misspecified models (e.g., 458.3563 for $\mathbf{M}_3$). In the mixed structure, weights are distributed more evenly (e.g., 0.4666, 0.4315, 0.1019 at $T = 10, n = 225$), with MA loss (7.8890) outperforming single models (e.g., 9.7469 for $\mathbf{M}_3$). Increasing T from 5 to 10 enhances precision (e.g., $\mathbf{M}_2$ loss drops from 3.6366 to 1.3029), and larger n reduces RMSE and loss, reinforcing the method's robustness and accuracy in spatial error modeling. The histograms of residuals in Fig. 4 for the mixed spatial structure scenario ($\mathbf{M}_2 + \mathbf{M}_3$) reveal an approximately normal distribution centered at zero with symmetry, suggesting that the model fits the data well. Furthermore, the scatter plots of actual versus fitted values exhibit a close alignment along the 45-degree line, with only slight deviations at the extremes, which underscores the predictive accuracy of the MA framework.



(a) $T = 5, n = 100$, Histogram of Residuals (b) $T = 5, n = 100$, Actual vs. Fitted Values (c) $T = 10, n = 100$, Histogram of Residuals (d) $T = 10, n = 100$, Actual vs. Fitted Values



(e) $T = 5, n = 225$, Histogram of Residuals (f) $T = 5, n = 225$, Actual vs. Fitted Values (g) $T = 10, n = 225$, Histogram of Residuals (h) $T = 10, n = 225$, Actual vs. Fitted Values

Fig. 4. Histogram of residuals and Actual vs. Fitted values for SEM: True matrices $\mathbf{M}_2 + \mathbf{M}_3$ ($T = 5$ and $T = 10$).

5.3. Comparative study: MMA vs. classical criteria

We compare the MA method (MMA) with C_p, AIC, and BIC using a simplified SARAR model: $y = \rho \mathbf{W} y + \mathbf{X}\boldsymbol{\beta} + u$, $u = \rho \mathbf{W} u + \varepsilon$, with $\rho \in \{0, 0.2, 0.5, 0.8\}$, true matrices $\mathbf{W}_1$ or $\mathbf{W}_3$, and candidates $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{W}_4$. The design follows Section 5.1.1, with $\mathbf{X} \sim \mathcal{N}(0, 1)$, $\varepsilon \sim \mathcal{N}(0, 0.1\mathbf{I})$, and response generated via $y = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\varepsilon)$. Parameters are estimated via QMLE, and MMA optimizes weights using quadratic programming with an adaptive penalty based on $\|\mathbf{X}\hat{\boldsymbol{\beta}}_{s,h} - \mathbf{X}\hat{\boldsymbol{\beta}}_{1,1}\|^2$.

Table 6 reveal significant spatial heterogeneity in the performance of model selection criteria, deeply influenced by the interaction of spatial dependence intensity (ρ), time dimension (T), and sample size (n). For the true model W_3 (Rook adjacency), under weak spatial dependence ($\rho = 0.2$), traditional criteria (C_p/AIC/BIC) achieve 100% correct selection at $T = 5$, while MMA assigns 69.6% weight to W_3 ($n = 100$), distributing the remainder to other candidates. Increasing n to 225 and T to 10 raises MMA's weight for W_3 to 78.8%, demonstrating synergistic enhancement from spatiotemporal information. In moderate dependence ($\rho = 0.5$), MMA attains 92.9% weight for W_3 ($n = 100$, $T = 5$), with misassignment reduced by 82% compared to $\rho = 0.2$. Temporal extension further concentrates weights to 95.6%, highlighting uncertainty compression. Under strong dependence ($\rho = 0.8$), all methods perform optimally, yet MMA exhibits minor dispersion (86.1% weight for W_3 at $n = 100$, $T = 5$), mitigated to 88.3% with $n = 225$. For the null model W_1 ($\rho = 0$), MMA achieves perfect identification (100% weight), whereas C_p selects W_1 with only 80.0% probability ($n = 100$, $T = 5$), and AIC/BIC show dispersed weights (e.g., 18.7% for W_1 under AIC). Temporal extension improves C_p's accuracy to 56.0% ($T = 10$), while MMA maintains 100% reliability. Spatiotemporal interaction analysis shows temporal extension ($T = 5 \rightarrow 10$) enhances MMA performance (for $\rho = 0.2$) more than cross-sectional expansion ($n = 100 \rightarrow 225$). Traditional criteria remain insensitive to n, reflecting variance insensitivity. MMA's superiority stems from dynamic penalties, weight smoothing for small-sample robustness, and efficient temporal information utilization. Overall, MMA outperforms traditional criteria in controlling overfitting, enhancing discriminative robustness.

5.4. Extension to irregular spatial weight matrices

In spatial econometrics, spatial units like administrative regions or economic zones often diverge from regular lattice structures. To assess our proposed model averaging framework's robustness in such scenarios, we extend Monte Carlo simulations using the SAR model, incorporating six spatial weight matrices: three regular (Zero, Queen, Rook) and three irregular (Inverse Distance Weighting (IDW), k-Nearest Neighbors (KNN), and Economic). The IDW matrix, built from random coordinates ($n = 100$), uses an exponential decay function $w_{ij} = \exp(-1.4 \cdot d_{ij}/100)$ based on scaled Euclidean distances, row-standardized with zero diagonals. The KNN matrix links each point to its four nearest neighbors with equal weights (1/4), while the Economic matrix, derived from random GDP values (0 to 100), employs $w_{ij} = 1/(1 + |GDP_i - GDP_j|)$, also row-standardized. The true spatial weight matrix $\mathbf{W}_{\text{true}}$ is a convex combination of the IDW and Economic matrices with a mixing parameter $w_{\text{weight}} = 0.5$, alongside parameters $\beta = [1.5, -2.0, 0.5]^T$, $\lambda = 0.5$, and $\sigma^2 = 0.01$, consistent with prior experiments in Section 5.1. This setup evaluates the framework's ability to handle mixed irregular spatial dependencies, including economic influences, in non-lattice settings.

Table 7 summarizes the performance across the four scenarios. For $T = 5$ and $n = 100$, the average loss is 0.4222, with RMSE values of 0.0146 for $\hat{\lambda}$ and 0.0056 for $\hat{\beta}$. Increasing the spatial units to $n = 225$ (with $T = 5$) reduces the average loss to 0.3124 and improves the RMSE values to 0.0130 for $\hat{\lambda}$ and 0.0036 for $\hat{\beta}$, highlighting the advantage of a larger cross-sectional sample in

Table 6
Performance comparison of different model selection criteria.

	$\rho = 0.2$	True **W** is $\mathbf{W}_3$ (T = 5)				True **W** is $\mathbf{W}_3$ (T = 10)			
		$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$
n = 100	C_p	0	0	1	0	0	0	1	0
	MMA	0.295	0.008	0.696	0.001	0.350	0	0.650	0
	AIC	0	0	1	0	0	0	1	0
	BIC	0	0	1	0	0	0	1	0
n = 225	C_p	0	0	1	0	0	0	1	0
	MMA	0.165	0.013	0.822	0.000	0.212	0.000	0.788	0.000
	AIC	0	0	1	0	0	0	1	0
	BIC	0	0	1	0	0	0	1	0
	$\rho = 0.5$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$
n = 100	C_p	0	0	1	0	0	0	1	0
	MMA	0.026	0.034	0.929	0.011	0.018	0.022	0.956	0.004
	AIC	0	0	1	0	0	0	1	0
	BIC	0	0	1	0	0	0	1	0
n = 225	C_p	0	0	1	0	0	0	1	0
	MMA	0.019	0.029	0.943	0.008	0.019	0.025	0.952	0.004
	AIC	0	0	1	0	0	0	1	0
	BIC	0	0	1	0	0	0	1	0
	$\rho = 0.8$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$
n = 100	C_p	0	0	1	0	0	0	1	0
	MMA	0.053	0.047	0.861	0.040	0.044	0.026	0.901	0.029
	AIC	0	0	1	0	0	0	1	0
	BIC	0	0	1	0	0	0	1	0
n = 225	C_p	0	0	1	0	0	0	1	0
	MMA	0.038	0.055	0.883	0.024	0.046	0.031	0.900	0.023
	AIC	0	0	1	0	0	0	1	0
	BIC	0	0	1	0	0	0	1	0
	$\rho = 0$	True **W** is $\mathbf{W}_1$ (T = 5)				True **W** is $\mathbf{W}_1$ (T = 10)			
		$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$
n = 100	C_p	0.800	0.120	0.060	0.020	0.560	0.220	0.120	0.100
	MMA	1.000	0	0	0	1.000	0	0	0
	AIC	0.187	0.279	0.284	0.250	0.158	0.301	0.289	0.252
	BIC	0.187	0.279	0.284	0.250	0.158	0.301	0.289	0.252
n = 225	C_p	0.680	0.120	0.060	0.140	0.640	0.120	0.060	0.180
	MMA	1.000	0	0	0	1.000	0.000	0.000	0.000
	AIC	0.166	0.265	0.259	0.310	0.166	0.270	0.264	0.300
	BIC	0.166	0.265	0.259	0.310	0.166	0.270	0.264	0.300

Table 7
Model averaging performance for SAR in irregular spatial weight matrices: True matrix $\mathbf{W}_4 + \mathbf{W}_5$.

	T = 5	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_5$	$\mathbf{W}_6$	MA
$n = 100$	RMSE of $\hat{\lambda}$	0.0000	0.4577	0.4807	0.0163	0.0146	0.4730	0.0146
	RMSE of $\hat{\boldsymbol{\beta}}$	0.0122	0.0115	0.0117	0.0054	0.0057	0.0113	0.0056
	Loss	28.8453	27.0758	27.9525	0.4842	0.4737	27.8998	0.4222
	MA weights	0.0357	0.0000	0.0000	0.4114	0.5529	0.0113	–
$n = 225$	RMSE of $\hat{\lambda}$	0.0000	0.4746	0.4867	0.0147	0.0125	0.4830	0.0130
	RMSE of $\hat{\boldsymbol{\beta}}$	0.0058	0.0059	0.0060	0.0037	0.0036	0.0055	0.0036
	Loss	32.0208	30.9034	31.4402	0.3007	0.3077	31.3055	0.3124
	MA weights	0.0279	0.0000	0.0000	0.4722	0.5000	0.0000	–
	T = 10	$\mathbf{W}_1$	$\mathbf{W}_2$	$\mathbf{W}_3$	$\mathbf{W}_4$	$\mathbf{W}_5$	$\mathbf{W}_6$	MA
$n = 100$	RMSE of $\hat{\lambda}$	0.0000	0.4560	0.4773	0.0087	0.0090	0.4690	0.0087
	RMSE of $\hat{\boldsymbol{\beta}}$	0.0098	0.0095	0.0095	0.0022	0.0022	0.0090	0.0021
	Loss	57.8474	53.9628	85.8446	0.6159	0.6150	55.4465	0.4713
	MA weights	0.0127	0.0000	0.0000	0.5640	0.4233	0.0000	–
$n = 225$	RMSE of $\hat{\lambda}$	0.0195	0.4709	0.4863	0.0053	0.0070	0.4813	0.0061
	RMSE of $\hat{\boldsymbol{\beta}}$	0.0063	0.0059	0.0062	0.0023	0.0023	0.0061	0.0023
	Loss	76.4969	73.3236	74.8490	0.6197	0.6664	74.6738	0.4954
	MA weights	0.0092	0.0000	0.0000	0.4703	0.5205	0.0000	–

enhancing estimation accuracy. Extending the time dimension to $T = 10$ with $n = 100$ increases the average loss slightly to 0.4713, but improves RMSE values to 0.0087 for $\hat{\lambda}$ and 0.0021 for $\hat{\beta}$, suggesting that a longer time period reduces estimation variability, particularly for regression coefficients. For $T = 10$ and $n = 225$, the average loss is 0.4954, with RMSE values of 0.0061 for $\hat{\lambda}$ and 0.0023 for $\hat{\beta}$, demonstrating the combined benefit of larger n and T in optimizing performance. The model averaging weights predominantly favor the true combination $\mathbf{W}_4 + \mathbf{W}_5$, with weights ranging from 0.9233 to 0.9908 across scenarios, indicating robust identification of the irregular spatial structure.

These results underscore the framework's adaptability to irregular spatial weight matrices, demonstrating its effectiveness in identifying and weighting complex spatial dependencies driven by distance and economic factors. The consistent improvement in RMSE and loss with increasing sample size highlights the practical utility of the approach for non-lattice spatial data.

5.5. Computational efficiency

The proposed model averaging framework achieves significant computational efficiency through three key mechanisms. First, the orthogonal transformation matrix $\mathbf{Q} = \mathbf{F}'_{T,T-1} \otimes \mathbf{I}_n$ eliminates individual fixed effects while reducing the temporal dimension from T to $T-1$, shrinking the parameter space from $n + k + 3$ to $k + 3$. This dimensional reduction avoids the variance estimation bias plim $\hat{\sigma}^{2d} = \sigma_0^2 \cdot \frac{T-1}{T}$ inherent in dummy methods when T is fixed, while ensuring consistent estimation and asymptotic normality without additional bias-correction steps, substantially lowering computational overhead for large n. (see the supplementary materials) Second, sparse matrix optimizations leverage the inherent structure of spatial weight matrices $\mathbf{W}_n$ and $\mathbf{M}_n$. Critical operations like log-determinant calculations for $|\mathbf{I}_n - \lambda \mathbf{W}_n|$ achieve $O(n)$ complexity via Chebyshev approximation or partial eigenvalue decomposition, bypassing $O(n^3)$ full matrix inversion. The Kronecker product formulation $(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda))\mathbf{Y}$ enables efficient block-sparse operations, minimizing memory and computational demands during quasi-maximum likelihood estimation across candidate models. Finally, the model averaging weight optimization $\hat{\mathbf{w}} = \arg\min_{\mathbf{w} \in \mathcal{H}} \hat{C}(\mathbf{w})$ is solved as a quadratic programming problem with linear constraints, efficiently implemented using established algorithms (e.g., MATLAB's `quadprog`). The criterion $\hat{C}(\mathbf{w})$ admits parallel computation of candidate-specific components $D_{s,h}$ and $\hat{\mathbf{P}}_{s,h}$, with interior-point methods ensuring scalability even for large $S \times H$ model sets, maintaining tractability for high-dimensional spatiotemporal analyses.

Future developments may involve the integration of stochastic optimization techniques for large-scale samples and the utilization of GPU acceleration for efficient sparse matrix computations, thereby broadening the framework's applicability to extensive spatiotemporal datasets without compromising computational feasibility.

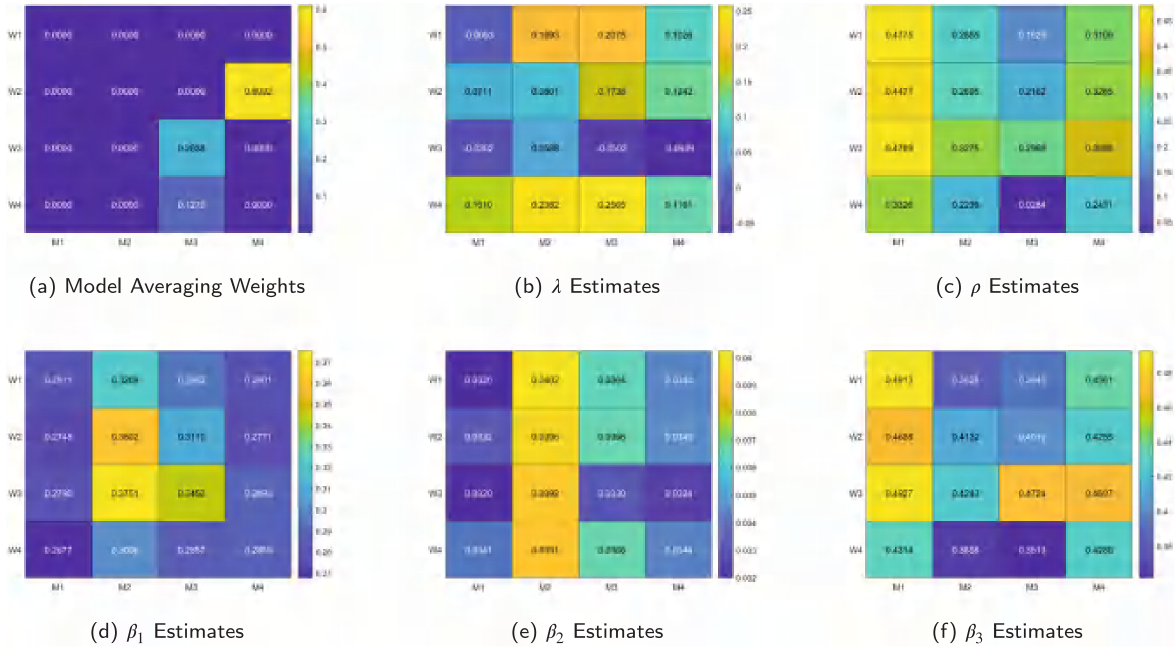
6. Empirical experiment in housing prices

To demonstrate the practical applicability of the proposed model averaging (MA) approach for spatial panel SARAR models, an empirical analysis of housing prices across 31 Chinese provinces from 2013 to 2022 is conducted, utilizing a spatial panel dataset with $31 \times 10 = 310$ observations to explore spatial dependence patterns driven by economic and geographical factors (Liu et al., 2018). Housing prices serve as a critical indicator of regional economic vitality, with elevated prices often signaling prosperity, population inflow, and employment opportunities, while depressed prices may indicate economic slowdown or population outflow. Analyzing spatial disparities in housing prices provides insights into regional economic health, resource allocation, industrial structure, and infrastructure development, informing policy decisions to address imbalances.

The dataset, sourced from the China Statistical Yearbook (https://www.stats.gov.cn/sj/ndsj/), includes average residential property prices per square meter as the dependent variable, with three key explanatory variables: per capita disposable income (reflecting purchasing power), population density (capturing urbanization), and infrastructure investment (measuring development). An orthogonal transformation $\mathbf{Q} = \mathbf{F}'_{T,T-1} \otimes \mathbf{I}_n$ eliminates individual fixed effects, reducing the time dimension from $T = 10$ to $T - 1 = 9$, yielding $31 \times 9 = 279$ observations for estimation. Four candidate spatial weight matrices are defined for both $\mathbf{W}_s$ and $\mathbf{M}_h$ $(s, h = 1, \ldots, 4)$, generating $4 \times 4 = 16$ model combinations: $\mathbf{W}_1, \mathbf{M}_1$ as a geographical adjacency matrix (1 for adjacent provinces, 0 otherwise, row-normalized); $\mathbf{W}_2, \mathbf{M}_2$ as an economic adjacency matrix; $\mathbf{W}_3, \mathbf{M}_3$ as a first-order distance band matrix (distance $\leq$300 km, row-normalized); and $\mathbf{W}_4, \mathbf{M}_4$ as an exponential decay matrix ($w_{ij} = \exp(-1.4 \times (d_{ij}/100))$, row-normalized. Keller and Shiue (2007) indicate that the specification with $\theta_d = -1.4$ effectively fits the data, determined through a constrained grid search based on likelihood. Model estimation employs QML estimators to compute parameters λ, ρ, β for each candidate model, followed by weighted averaging using a Mallows-type criterion.

The results presented in Fig. 5, reveal that the MA approach assigns the highest weight (0.6092) to the model combining the economic adjacency matrix ($\mathbf{W}_2$) for the dependent variable and the exponential decay matrix ($\mathbf{M}_4$) for the error term, indicating that spatial dependence in Chinese provincial housing prices is primarily driven by economic interactions, with error autocorrelation following an exponential decay pattern. Additional notable weights include 0.2638 for the combination of $\mathbf{W}_3$ and $\mathbf{M}_3$, and 0.1270 for $\mathbf{W}_4$ and $\mathbf{M}_3$. The $\boldsymbol{\beta}$ coefficients, shown in Figs. 5(d)–5(f), suggest that per capita disposable income exerts the strongest influence on housing prices, with estimates ranging from 0.2748 to 0.3751 across models, followed by infrastructure investment (ranging from 0.3613 to 0.4927), while population density has a smaller effect (ranging from −0.0320 to 0.0402). These findings are consistent with econometric theory linking income and urbanization to real estate demand. Parameter estimates for λ (Fig. 5(b)) range from −0.1610 to 0.2585 across models, indicating moderate spatial lag dependence, while ρ estimates (Fig. 5(c)) vary from 0.0284 to 0.4789, reflecting weaker but significant error autocorrelation.

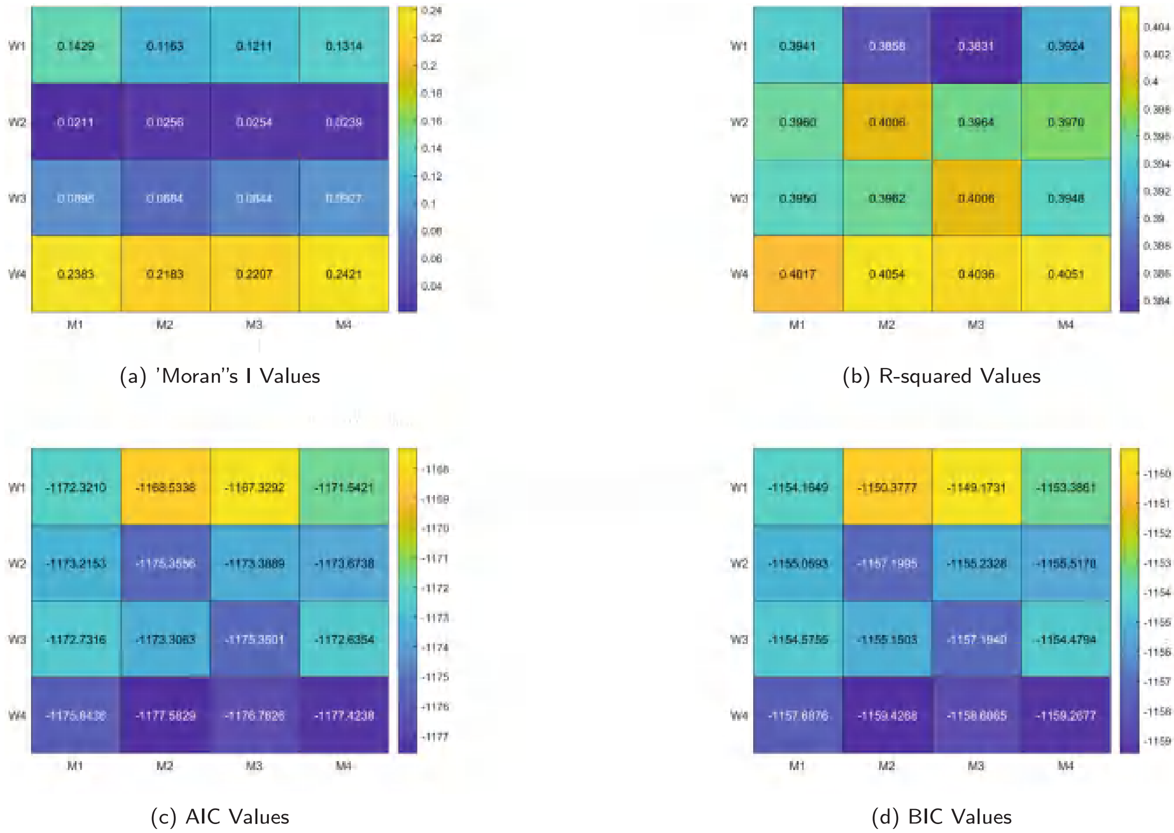
The Diagnostic metrics (Fig. 6) validate model performance: Moran's I (0.0211–0.2421) indicates minimal residual autocorrelation, particularly in models with economic/exponential decay matrices; R^2 (0.3831–0.4054) shows reasonable explanatory power;



(a) Model Averaging Weights (b) λ Estimates (c) ρ Estimates

(d) β_1 Estimates (e) β_2 Estimates (f) β_3 Estimates

Fig. 5. Empirical experiment results.



(a) 'Moran''s I Values (b) R-squared Values

(c) AIC Values (d) BIC Values

Fig. 6. Empirical experiment results diagnostic.

Table 8
Model averaging estimates and diagnostics.

Parameter	Estimate
λ_{MA}	0.1250
ρ_{MA}	0.2808
β_{MA1} (Income)	0.2962
β_{MA2} (Density)	0.0344
β_{MA3} (Investment)	0.4297
σ^2_{MA}	0.0122
Diagnostic	Value
R-squared	0.4022
AIC	−1176.1029
BIC	−1157.9469
Moran's I	0.0723

Table 9
Spatial effects from model averaging.

Variable	Direct effect	Indirect effect	Total effect
Per capita income	0.2977	0.0346	0.4319
Population density	0.0336	0.0043	0.0487
Infrastructure investment	0.3312	0.0388	0.4804

and lower AIC (−1177.58 to −1167.33)/BIC (−1159.43 to −1149.16) values favor exponential decay specifications. The diagnostic metrics, presented in Fig. 6, provide further validation of the model performance. Moran's I values, which measure residual spatial autocorrelation, range from 0.0211 to 0.2421 across the candidate models, with lower values in models incorporating economic and exponential decay matrices, suggesting effective capture of spatial dependence and minimal remaining autocorrelation in the residuals. The R-squared values, ranging from 0.3831 to 0.4054, indicate that the models explain approximately 38% to 41% of the variance in housing prices, reflecting reasonable explanatory power given the complexity of spatial panel data and the inclusion of fixed effects. The AIC and BIC values, ranging from −1177.5829 to −1167.3292 for AIC and −1159.4268 to −1149.1649 for BIC, serve as criteria for model comparison, where lower values denote better balance between fit and complexity; models with exponential decay matrices tend to have lower (better) scores.

Table 8 summarizes MA results. The spatial lag parameter ($\lambda_{\mathrm{MA}} = 0.1250$) indicates moderate positive dependence: a 1% neighbor price increase raises local prices by 0.1250%. Error autocorrelation ($\rho_{\mathrm{MA}} = 0.2808$) suggests spatial factors captured by exponential decay. Coefficients confirm income ($\beta_{\mathrm{MA1}} = 0.2962$) and investment ($\beta_{\mathrm{MA3}} = 0.4297$) as dominant drivers, with density ($\beta_{\mathrm{MA2}} = 0.0344$) having marginal impact. The diagnostics reveal a R-squared of 0.4022, reflecting a solid fit ; AIC (−1176.10) and BIC (−1157.94) values support model parsimony; and Moran's I (0.0723) confirms minimal residual spatial patterns, validating the MA framework's effectiveness in handling model uncertainty. In summary, Table 8 consolidates the averaged parameters and diagnostics.

Spatial effect decomposition Table 9 reveals: income increases exhibit local dominance (direct: 0.2962) with modest spillovers (indirect: 0.0312; total: 0.3274); infrastructure investment generates strong local (0.4401) and regional spillovers (0.0463), yielding the highest total effect (0.4864); density effects remain limited (total: 0.0379). These findings align with model diagnostics and weight distributions.

The empirical analysis demonstrates how MA mitigates spatial structure uncertainty, distinguishes between dependent variable and error dependence, and enhances robustness through Mallows weighting. Results indicate economic adjacency and exponential decay matrices best capture China's provincial housing price dynamics, with income and infrastructure as key determinants. Policy implications include promoting economic integration for market stability and incorporating spatial error structures in infrastructure planning. This confirms the MA framework's utility for spatial panel analysis in real estate economics.

7. Conclusions and prospects

This paper proposes a novel model averaging framework for spatial panel SARAR models, significantly advancing spatial econometric methodology by addressing uncertainty in spatial dependence structures. It integrates multiple spatial weight matrices, $\mathbf{W}_s$ and $\mathbf{M}_h$, for the dependent variable and error terms, respectively, while employing orthogonal transformations to eliminate individual fixed effects. The proposed adaptive Mallows-type criterion dynamically adjusts to the presence or absence of spatial effects, enhancing robustness and predictive accuracy. Monte Carlo simulations validate its superior performance across scenarios with no, single, or mixed spatial dependencies. An empirical analysis of Chinese provincial housing prices from 2013 to 2022 confirms the framework's practical utility, identifying economic adjacency and exponential decay error structures as key drivers of spatial dependence. Major contributions include extending the Mallows framework to panel data, providing a systematic approach to handling multiple weight matrices, and ensuring computational efficiency for large-scale applications. This framework offers a robust tool for spatiotemporal data analysis in regional economics, environmental science, and urban studies.

Future research directions promise to further refine this framework's applicability. Extending the model averaging approach to dynamic spatial panel models could capture temporal lags and feedback effects, enhancing its relevance for evolving spatial systems. Integrating machine learning techniques, such as neural networks or ensemble methods, may improve the selection and weighting of spatial structures in high-dimensional settings. Exploring performance under non-Gaussian error distributions or heteroskedastic conditions could broaden robustness to complex real-world data. Additionally, applying the method to diverse empirical contexts, such as global trade networks or environmental pollution diffusion, could yield new insights into spatial interactions and inform policy design. These advancements position the proposed framework as a cornerstone for innovative spatial econometric analysis, with potential to address emerging challenges in mathematical modeling of spatiotemporal phenomena.

CRediT authorship contribution statement

Aibing Ji: Supervision, Conceptualization, Resources, Funding Acquisition. **Jingxuan Li:** Writing – original draft, Conceptualization, Software, Visualization. **Qingqing Li:** Writing – review & editing, Conceptualization, Data Curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors intend to express their gratitude for the financial support from Operations Research and Management Innovation Team of Hebei University (IT2023C02), and the Innovation Capacity Enhancement Program-Science and Technology Platform Project (22567623H).

Appendix A

A.1. Quasi-maximum likelihood function and derivatives

The log-likelihood function for the spatial panel SARAR model is defined to facilitate parameter estimation for $\theta = (\lambda, \rho, \boldsymbol{\beta}, \sigma^2)$. The function is given by:

$$\ln L(\theta) = -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + \ln\left|\mathbf{I}_{T-1}\otimes\mathbf{S}_n(\lambda)\right| + \ln\left|\mathbf{I}_{T-1}\otimes\mathbf{R}_n(\rho)\right| - \frac{1}{2\sigma^2}\mathbf{V}^\top(\theta)\mathbf{V}(\theta),$$

where the transformed error vector is $\mathbf{V}(\theta) = \left(\mathbf{I}_{T-1}\otimes\mathbf{R}_n(\rho)\right)\left[\left(\mathbf{I}_{T-1}\otimes\mathbf{S}_n(\lambda)\right)\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right]$. Parameters $\hat{\rho}_{s,h}$, $\hat{\lambda}_{s,h}$, $\hat{\boldsymbol{\beta}}_{s,h}$, and $\hat{\sigma}^2_{s,h}$ are obtained by maximizing $\ln L(\theta)$. Below, we derive the necessary partial derivatives for QML estimators and the Mallows criterion.

Partial derivative with respect to ρ

To compute the first-order partial derivative of the log-likelihood with respect to ρ, we express:

$$\frac{\partial \ln L(\theta)}{\partial\rho} = \frac{\partial}{\partial\rho}\ln\left|\mathbf{I}_{T-1}\otimes\mathbf{R}_n(\rho)\right| - \frac{1}{2\sigma^2}\frac{\partial}{\partial\rho}\left(\mathbf{V}^\top(\theta)\mathbf{V}(\theta)\right).$$

The first term involves the determinant $\ln\left|\mathbf{I}_{T-1}\otimes\mathbf{R}_n(\rho)\right| = (T-1)\ln\left|\mathbf{R}_n(\rho)\right|$, with derivative $\frac{\partial \ln|\mathbf{R}_n(\rho)|}{\partial\rho} = \operatorname{tr}\left(\mathbf{R}_n(\rho)^{-1}\frac{\partial\mathbf{R}_n(\rho)}{\partial\rho}\right)$. Since $\mathbf{R}_n(\rho) = \mathbf{I}_n - \rho\mathbf{M}_n$, we have: $\frac{\partial\mathbf{R}_n(\rho)}{\partial\rho} = -\mathbf{M}_n$, yielding $\frac{\partial \ln|\mathbf{I}_{T-1}\otimes\mathbf{R}_n(\rho)|}{\partial\rho} = -(T-1)\operatorname{tr}\left(\mathbf{R}_n(\rho)^{-1}\mathbf{M}_n\right)$.

For the second term, define $\mathbf{Z} = \left(\mathbf{I}_{T-1}\otimes\mathbf{S}_n(\lambda)\right)\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$, so:

$$\mathbf{V}(\theta) = \left(\mathbf{I}_{T-1}\otimes\mathbf{R}_n(\rho)\right)\mathbf{Z}, \quad \frac{\partial\mathbf{V}(\theta)}{\partial\rho} = -\left(\mathbf{I}_{T-1}\otimes\mathbf{M}_n\right)\mathbf{Z}.$$

Thus:

$$\frac{\partial\left(\mathbf{V}^\top(\theta)\mathbf{V}(\theta)\right)}{\partial\rho} = 2\mathbf{V}^\top(\theta)\frac{\partial\mathbf{V}(\theta)}{\partial\rho} = -2\mathbf{Z}^\top\left(\mathbf{I}_{T-1}\otimes\mathbf{R}_n(\rho)\right)^\top\left(\mathbf{I}_{T-1}\otimes\mathbf{M}_n\right)\mathbf{Z}.$$

Combining these, the first-order derivative is:

$$\frac{\partial \ln L(\theta)}{\partial\rho} = -(T-1)\operatorname{tr}\left(\mathbf{R}_n(\rho)^{-1}\mathbf{M}_n\right) + \frac{1}{\sigma^2}\mathbf{Z}^\top\left(\mathbf{I}_{T-1}\otimes\mathbf{R}_n(\rho)\right)^\top\left(\mathbf{I}_{T-1}\otimes\mathbf{M}_n\right)\mathbf{Z}.$$

To derive $\frac{\partial\hat{\rho}_{s,h}}{\partial\mathbf{Y}^\top}$, we apply the implicit function theorem, setting $\frac{\partial \ln L(\theta)}{\partial\rho} = 0$ and differentiating with respect to $\mathbf{Y}^\top$:

$$\frac{\partial}{\partial\mathbf{Y}^\top}\left(\frac{\partial \ln L(\theta)}{\partial\rho}\right) = \frac{\partial\hat{\rho}}{\partial\mathbf{Y}^\top}\frac{\partial^2 \ln L(\theta)}{\partial\rho^2} + \frac{\partial}{\partial\mathbf{Y}^\top}\Big(-(T-1)\operatorname{tr}\left(\mathbf{R}_n(\rho)^{-1}\mathbf{M}_n\right) + \frac{1}{\sigma^2}\mathbf{Z}^\top\left(\mathbf{I}_{T-1}\otimes\mathbf{R}_n(\rho)\right)^\top\left(\mathbf{I}_{T-1}\otimes\mathbf{M}_n\right)\mathbf{Z}\Big) = 0.$$

Since $\frac{\partial \mathbf{Z}}{\partial \mathbf{Y}^\top} = \left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda)\right)$, we compute:

$$\frac{\partial}{\partial \mathbf{Y}^\top}\left[\mathbf{Z}^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho)\right)^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{M}_n\right)\mathbf{Z}\right] = 2\left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda)\right)^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho)\right)^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{M}_n\right)\mathbf{Z}.$$

Thus:

$$\frac{\partial \hat{\rho}_{s,h}}{\partial \mathbf{Y}^\top} = -\left(\frac{\partial^2 \ln L(\theta)}{\partial \rho^2}\right)^{-1} \frac{2}{\sigma^2}\left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\lambda)\right)^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho)\right)^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{M}_n\right)\mathbf{Z},$$

where $(\frac{\partial^2 \ln L(\theta)}{\partial \rho^2})$ is the scalar information matrix term, not further expanded here.

Partial derivative with respect to (λ)

Similarly, to compute $(\frac{\partial \hat{\lambda}_{s,h}}{\partial \mathbf{Y}^\top})$, we apply the implicit function theorem to the first-order condition $(\frac{\partial \ln L(\theta)}{\partial \lambda} = 0)$. The log-likelihood derivative with respect to λ involves terms analogous to those for ρ, leading to:

$$\frac{\partial \hat{\lambda}_{s,h}}{\partial \mathbf{Y}^\top} = -\left(\frac{\partial^2 \ln L(\theta)}{\partial \lambda^2}\right)^{-1} \frac{2}{\sigma^2}\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\rho)\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{W}_n\right)\mathbf{V}(\theta),$$

where $\mathbf{V}(\theta)$ is the transformed error vector defined above, and $(\frac{\partial^2 \ln L(\theta)}{\partial \lambda^2})$ is the corresponding information matrix term.

Derivatives of the projection matrix

The projection matrix for model (s, h) is:

$$\tilde{\mathbf{P}}_{s,h} = \left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right)^{-1}\mathbf{X}\Big(\mathbf{X}^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})^\top\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right)\mathbf{X}\Big)^{-1} \\ \times \mathbf{X}^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})^\top\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right).$$

Define:

$$\mathbf{A} = \left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right)^{-1}, \quad \mathbf{B} = \mathbf{X}\left(\mathbf{X}^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})^\top\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right)\mathbf{X}\right)^{-1}\mathbf{X}^\top,$$

$$\mathbf{C} = \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})^\top\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right).$$

Thus, $\tilde{\mathbf{P}}_{s,h} = \mathbf{ABC}$. The partial derivative with respect to $\hat{\rho}_{s,h}$ is:

$$\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\rho}_{s,h}} = \mathbf{A}\frac{\partial \mathbf{B}}{\partial \hat{\rho}_{s,h}}\mathbf{C} + \mathbf{AB}\frac{\partial \mathbf{C}}{\partial \hat{\rho}_{s,h}},$$

where:

$$\mathbf{D} = \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})^\top\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right),$$

$$\frac{\partial \mathbf{D}}{\partial \hat{\rho}_{s,h}} = -\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})^\top\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{M}_n\right) - \left(\mathbf{I}_{T-1} \otimes \mathbf{M}_n\right)^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right),$$

$$\frac{\partial \mathbf{B}}{\partial \hat{\rho}_{s,h}} = \mathbf{X}\left(\mathbf{X}^\top \mathbf{D}\mathbf{X}\right)^{-1}\left(\mathbf{X}^\top \frac{\partial \mathbf{D}}{\partial \hat{\rho}_{s,h}}\mathbf{X}\right)\left(\mathbf{X}^\top \mathbf{D}\mathbf{X}\right)^{-1}\mathbf{X}^\top,$$

$$\frac{\partial \mathbf{C}}{\partial \hat{\rho}_{s,h}} = \left[-\left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})^\top\right)\left(\mathbf{I}_{T-1} \otimes \mathbf{M}_n\right) - \left(\mathbf{I}_{T-1} \otimes \mathbf{M}_n\right)^\top \left(\mathbf{I}_{T-1} \otimes \mathbf{R}_n(\hat{\rho}_{s,h})\right)\right]\left(\mathbf{I}_{T-1} \otimes \mathbf{S}_n(\hat{\lambda}_{s,h})\right).$$

Hence:

$$\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\rho}_{s,h}}\mathbf{Y} = \left(\mathbf{A}\frac{\partial \mathbf{B}}{\partial \hat{\rho}_{s,h}}\mathbf{C} + \mathbf{AB}\frac{\partial \mathbf{C}}{\partial \hat{\rho}_{s,h}}\right)\mathbf{Y}.$$

For $\hat{\lambda}_{s,h}$, the derivative is:

$$\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\lambda}_{s,h}}\mathbf{Y} = \left(\frac{\partial \mathbf{A}}{\partial \hat{\lambda}_{s,h}}\mathbf{BC} + \mathbf{AB}\frac{\partial \mathbf{C}}{\partial \hat{\lambda}_{s,h}}\right)\mathbf{Y},$$

where $\frac{\partial \mathbf{A}}{\partial \hat{\lambda}_{s,h}}$ and $\frac{\partial \mathbf{C}}{\partial \hat{\lambda}_{s,h}}$ follow from differentiating $\mathbf{S}_n(\hat{\lambda}_{s,h})$.

A.2. Unbiased estimation of the Mallows criterion

To establish the unbiasedness of the Mallows-type criterion $C(\mathbf{w})$, we derive the expected risk using Stein's identity, ensuring the complexity penalty terms align with the model's degrees of freedom. Define the transformed response $(\mathbf{z} = \boldsymbol{\Omega}^{1/2}\mathbf{Y})$ (see Greven and Kneib (2010)), where $(\boldsymbol{\Omega})$ is the covariance matrix. Stein's identity yields:

$$\mathbb{E}\left\{\left(\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y}\right)^\top (\mathbf{Y} - \boldsymbol{\mu})\right\} = \mathbb{E}\left\{\left(\boldsymbol{\Omega}^{1/2}\tilde{\mathbf{P}}(\mathbf{w})\boldsymbol{\Omega}^{1/2}\mathbf{z}\right)^\top \left(\mathbf{z} - \boldsymbol{\Omega}^{-1/2}\boldsymbol{\mu}\right)\right\}$$

Table 10
Notation summary for spatial panel SARAR framework.

Symbol	Description
n	Number of spatial units (e.g., provinces, grid cells)
T	Number of time periods
Y_{nt}	$n \times 1$ vector of dependent variables at time t
X_{nt}	$n \times k$ matrix of explanatory variables at time t
$\mathbf{W}_n, \mathbf{M}_n$	$n \times n$ spatial weight matrices for dependent variable and errors
λ	Spatial autoregressive parameter for dependent variable
ρ	Spatial autoregressive parameter for error term
$\boldsymbol{\beta}$	$k \times 1$ vector of regression coefficients
c_{n0}	$n \times 1$ vector of individual fixed effects
U_{nt}	Spatially autocorrelated error term
V_{nt}	i.i.d. innovation term $\sim (0, \sigma_0^2)$
Symbol	Description
$\mathbf{Y}^*, \mathbf{X}^*, \mathbf{V}^*$	Stacked panel data vectors/matrices ($NT \times 1$ or $NT \times K$)
Q	Orthogonal transformation matrix ($F_{T,T-1} \otimes I_n$)
$F_{T,T-1}$	Fixed effects elimination matrix ($T \times (T-1)$)
$\mathbf{S}_n(\lambda)$	Spatial filter $I_n - \lambda \mathbf{W}_n$
$\mathbf{R}_n(\rho)$	Error spatial filter $I_n - \rho \mathbf{M}_n$
$\boldsymbol{\mu}$	Conditional expectation $E(\mathbf{Y})$ (Eq. (3))
$\boldsymbol{\Omega}$	Variance–covariance matrix of $\mathbf{Y}$
Symbol	Description
$\mathcal{W}, \mathcal{M}$	Candidate sets for $\mathbf{W}$ and $\mathbf{M}$ matrices
S, H	Number of candidate matrices for $\mathcal{W}$ and $\mathcal{M}$
$\theta_{s,h}$	Parameter vector $(\lambda_{s,h}, \rho_{s,h}, \boldsymbol{\beta}_{s,h})$ for model (s,h)
$\hat{\boldsymbol{\mu}}_{s,h}$	Estimated mean for candidate model (s,h) (Eq. (6))
$\tilde{\mathbf{P}}_{s,h}$	Projection matrix for model (s,h) (Eq. (8))
$w_{s,h}$	Weight assigned to model (s,h)
$\mathbf{w}$	Weight vector $(w_{1,1}, \ldots, w_{S,H})^\top$
$\mathcal{H}$	Simplex constraint set $\{\mathbf{w} \in [0,1]^{SH} : \sum_s \sum_h w_{s,h} = 1\}$
$\hat{\boldsymbol{\mu}}(\mathbf{w})$	Model-averaged estimator (Eq. (13))
$C(\mathbf{w})$	Mallows-type criterion function (Eq. (14))
$\mathcal{D}_{s,h}$	Model complexity measure
$\text{pen}_{s,h}$	Adaptive penalty term

$$= \mathbb{E}\left\{\text{trace}\left(\frac{\partial\left(\boldsymbol{\Omega}^{1/2}\tilde{\mathbf{P}}(\mathbf{w})\boldsymbol{\Omega}^{1/2}\mathbf{z}\right)}{\partial \mathbf{z}^\top}\right)\right\}$$

$$= \mathbb{E}\left\{\text{trace}\left(\boldsymbol{\Omega}^{1/2}\tilde{\mathbf{P}}(\mathbf{w})\boldsymbol{\Omega}^{1/2}\right)\right\} + \mathbb{E}\left[\text{trace}\left(\frac{\partial\left(\boldsymbol{\Omega}^{1/2}\tilde{\mathbf{P}}(\mathbf{w})\boldsymbol{\Omega}^{1/2}\mathbf{z}\right)}{\partial \hat{\rho}_{s,h}} \cdot \frac{\partial \hat{\rho}_{s,h}}{\partial \mathbf{z}^\top}\right)\right.$$

$$\left. + \text{trace}\left(\frac{\partial\left(\boldsymbol{\Omega}^{1/2}\tilde{\mathbf{P}}(\mathbf{w})\boldsymbol{\Omega}^{1/2}\mathbf{z}\right)}{\partial \hat{\lambda}_{s,h}} \cdot \frac{\partial \hat{\lambda}_{s,h}}{\partial \mathbf{z}^\top}\right)\right]$$

$$= \mathbb{E}\left[\sum_{s=1}^{S}\sum_{h=1}^{H} w_{s,h}\left\{\text{trace}\left(\tilde{\mathbf{P}}_{s,h}\boldsymbol{\Omega}\right) + \frac{\partial \hat{\rho}_{s,h}}{\partial \mathbf{Y}^\top}\boldsymbol{\Omega}\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\rho}_{s,h}}\mathbf{Y} + \frac{\partial \hat{\lambda}_{s,h}}{\partial \mathbf{Y}^\top}\boldsymbol{\Omega}\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\lambda}_{s,h}}\mathbf{Y}\right\}\right].$$

The risk is:

$$R(\mathbf{w}) = \mathbb{E}\left\|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \boldsymbol{\mu}\right\|^2 = \mathbb{E}\left\|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y} + \mathbf{Y} - \boldsymbol{\mu}\right\|^2 = \mathbb{E}\left\|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\right\|^2 + 2\mathbb{E}\left[\left(\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y}\right)^\top(\mathbf{Y} - \boldsymbol{\mu})\right] + \text{tr}(\boldsymbol{\Omega}).$$

Substituting the expectation derived above, we obtain:

$$R(\mathbf{w}) = \mathbb{E}\left\|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\right\|^2 + 2\mathbb{E}\left[\sum_{s=1}^{S}\sum_{h=1}^{H} w_{s,h}\left\{\text{tr}\left(\tilde{\mathbf{P}}_{s,h}\boldsymbol{\Omega}\right) + \frac{\partial \hat{\rho}_{s,h}}{\partial \mathbf{Y}^\top}\boldsymbol{\Omega}\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\rho}_{s,h}}\mathbf{Y} + \frac{\partial \hat{\lambda}_{s,h}}{\partial \mathbf{Y}^\top}\boldsymbol{\Omega}\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\lambda}_{s,h}}\mathbf{Y}\right\}\right] + \text{tr}(\boldsymbol{\Omega}),$$

confirming that $C(\mathbf{w}) = \|\tilde{\mathbf{P}}(\mathbf{w})\mathbf{Y} - \mathbf{Y}\|^2 + 2\sum_{s=1}^{S}\sum_{h=1}^{H} w_{s,h}\left\{\text{tr}\left(\tilde{\mathbf{P}}_{s,h}\boldsymbol{\Omega}\right) + \frac{\partial \hat{\rho}_{s,h}}{\partial \mathbf{Y}^\top}\boldsymbol{\Omega}\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\rho}_{s,h}}\mathbf{Y} + \frac{\partial \hat{\lambda}_{s,h}}{\partial \mathbf{Y}^\top}\boldsymbol{\Omega}\frac{\partial \tilde{\mathbf{P}}_{s,h}}{\partial \hat{\lambda}_{s,h}}\mathbf{Y}\right\}$ is an unbiased estimator of the risk, adjusted by $\text{tr}(\boldsymbol{\Omega})$.

Appendix B

B.1. Glossary of key symbols

See Table 10.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spasta.2025.100931.

References

Akaike, H., 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika 60 (2), 255–265.
Anselin, L., 1988. Spatial Econometrics: Methods and Models. Springer Netherlands.
Ben Abdallah, N., Dabbou, H., Gallali, M.I., Hathroubi, S., 2025. Stock returns' co-movement: A spatial model with convex combination of connectivity matrices. Risks 13 (6), 110.
Chen, H., Hu, Y., Tang, B., 2025. Traffic density estimation of urban road networks based on functional spatial autoregressive model averaging. Transp. Res. Rec. 2679 (7), 756–772.
Efron, B., 2004. The estimation of prediction error: Covariance penalties and cross-validation. J. Amer. Statist. Assoc. 99 (467), 619–632.
Greven, S., Kneib, T., 2010. On the behaviour of marginal and conditional AIC in linear mixed models. Biometrika 97 (4), 773–789.
Hansen, B.E., 2007. Least squares model averaging. Econometrica 75 (4), 1175–1189.
Jin, F., Lee, L., 2020. Asymptotically efficient root estimators for spatial autoregressive models with spatial autoregressive disturbances. Econom. Lett. 194, 109343.
Kelejian, H.H., 2008. A spatial J-test for model specification against a single or a set of non-nested alternatives. Lett. Spat. Resour. Sci. 1, 3–11.
Kelejian, H.H., Piras, G., 2011. An extension of Kelejian's J-test for non-nested spatial models. Reg. Sci. Urban Econ. 41 (3), 281–292.
Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. J. Econometrics 157 (1), 53–67.
Keller, W., Shiue, C.H., 2007. The origin of spatial interaction. J. Econometrics 140 (1), 304–332.
Lee, L., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. J. Econometrics 154 (2), 165–185.
LeSage, J.P., Pace, R.K., 2009. Introduction to Spatial Econometrics. CRC Press, Taylor and Francis Group, Boca Raton.
LeSage, J.P., Parent, O., 2007. Bayesian model averaging for spatial econometric models. Geogr. Anal. 39 (3), 241–267.
Li, K.-C., 1987. Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. Ann. Statist. 15 (3), 958–975.
Li, Q., Zheng, R., Ji, A., Ma, H., 2025. Fixed effects spatial panel interval-valued autoregressive models and applications. Spat. Stat. 65, 100875.
Liang, H., Zou, G., Wan, A.T., Zhang, X., 2011. Optimal weight choice for frequentist model average estimators. J. Amer. Statist. Assoc. 106 (495), 1053–1066.
Liao, J., Zou, G., Gao, Y., Zhang, X., 2021. Model averaging prediction for time series models with a diverging number of parameters. J. Econometrics 223 (1), 190–221.
Lin, T.-C., Liu, C.-A., 2025. Model averaging prediction for possibly nonstationary autoregressions. J. Econometrics 249, 105994.
Liu, F., Liu, C., Ren, H., 2018. Urban housing price fluctuations and regional systemic financial risks: Panel spatial economic models in Jiangsu, China. Sustain. 10 (10), 3452.
Liu, Q., Okui, R., 2013. Heteroscedasticity-robust Cp model averaging. Econom. J. 16 (3), 463–472.
Nguyen, M., Veraart, A.E., 2017. Modelling spatial heteroskedasticity by volatility modulated moving averages. Spat. Stat. 20, 148–190.
Piras, G., Sarrias, M., 2025. A J-test for spatial autoregressive binary models. Spat. Stat. 67, 100903.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.
Song, Y., Cibin, A., 2024. Optimizing spatial weight matrices in spatial econometrics: A graph-theoretic approach based on shortest path algorithms a New York city application of crime and economic indicators. Int. Rev. Spat. Plan. Sustain. Dev. 12 (2), 181–200.
Stein, C.M., 1981. Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9 (6), 1135–1151.
Su, B., Zhu, F.-k., Huang, J., 2023. Bayesian estimation and model selection for the spatiotemporal autoregressive model with autoregressive conditional heteroscedasticity errors. Acta Math. Appl. Sin. Engl. Ser. 39 (4), 972–989.
Tian, L., Wei, C., Wu, M., 2025. Statistical inference of partially linear time-varying coefficients spatial autoregressive panel data model. Spat. Stat. 66, 100887.
Wan, A.T., Zhang, X., Zou, G., 2010. Least squares model averaging by Mallows criterion. J. Econometrics 156 (2), 277–283.
Zhang, X., Yu, J., 2018. Spatial weights matrix selection and model averaging for spatial autoregressive models. J. Econometrics 203 (1), 1–18.
Zhang, X., Zou, G., Liang, H., 2014. Model averaging and weight choice in linear mixed-effects models. Biometrika 101 (1), 205–218.
Zou, H., Hastie, T., Tibshirani, R., 2007. On the degrees of freedom of the lasso. Ann. Statist. 35 (5), 2173–2192.
Zou, C., Li, X., Li, X., Liang, H., 2025. Asymptotic optimality of generalized cross validation and regularized Mallows model averaging. Statist. Probab. Lett. 222, 110406.