Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue on Recent Advances in Bayesian Statistics and Machine Learning
Volume 24, No. 2, 2026 (New Series), pp 1–21
http://www.ssca.org.in/journal

Bayesian Predictive Inference for Multiple Series with Correlated Spatial Priors on Autoregressive Parameters

Hammed A. Olayinka and Balgobin Nandram
Department of Mathematical Sciences
Worcester Polytechnic Institute, Worcester, MA 01609

Received: 14 November 2025; Revised: 14 April 2026; Accepted: 16 April 2026

Abstract

We present a Bayesian multiple series model for panel data that improves forecasting by borrowing information across neighboring areas. Each unit is modeled with a simple univariate autoregressive likelihood, while spatial dependence is introduced entirely through a conditional autoregressive prior placed on unit-level intercept and autoregressive coefficient deviations. Inference is conducted using a collapsed Gibbs sampler, which reduces posterior dependence, improves mixing and effective sample size, and preserves interpretable unit-level dynamics. This design avoids the over-parameterization of multivariate formulations while enabling spatially targeted pooling that respects geographic adjacency. We derive stable posterior updates under standard positive-definiteness conditions for the conditional autoregressive precision. Empirically, we study an annual panel of average hourly earnings for production employees across selected California metropolitan statistical areas and compare the proposed model to independent single-series autoregressive models and a pooled multiple time-series model without spatial structure as benchmarks. The proposed approach delivers consistent gains in point accuracy and interval calibration, providing a flexible and interpretable framework for forecasting and small-area estimation in regional economic analysis and related domains. All R codes available at: `https://github.com/haolayinka/MTS-CAR-Model`.

Key words: Gibbs sampler; Hierarchical Bayesian model; Markov-chain Monte Carlo algorithm; Panel forecasting; Pooling time series; Small area estimation.

AMS Subject Classifications: 62M10, 62M30, 62F15, 65C40

1. Introduction

In panel settings, three pragmatic Bayesian strategies are common. First, fit each series independently (*e.g.* Autoregressive of order p ($AR(p)$) per unit), which preserves interpretability but ignores recurring cross-unit patterns and can overfit noise when number of time (T) is small. Second, adopt fully multivariate specifications such as Bayesian

Corresponding Author: Hammed A. Olayinka
Email: haolayinka@wpi.edu

vector autoregressive, which explicitly model cross-sectional dependence but can be over-parameterized for moderate number of areas (ℓ) and modest T, require careful prior design to avoid over-shrinkage, and often entail substantial computation as the dimension grows. Third, pool information through shared hyperparameters or exchangeable hierarchical priors across units (*e.g.*, Nandram and Petruccelli (1997)), which stabilizes estimation yet treats units as roughly interchangeable and still does not exploit that nearby areas tend to move together.

Sims (1980) positioned vector autoregressive (VAR) as flexible, data-driven models for multiple time series, but their parameter count grows rapidly with the number of variables and lags, leading to overfitting in small panels and motivating shrinkage or alternative designs - here, adjacency-aware pooling via conditional autoregressive (CAR) priors on univariate autoregressive (AR) likelihoods. A central distinction is between contemporaneous covariance and spatial adjacency. Allowing a non-diagonal cross-sectional covariance in the likelihood captures same-time correlation, but it does not encode where borrowing should occur. In many regional applications, dependence decays with distance and is strongest along shared borders or network links. Models that omit adjacency risk "global" shrinkage - useful on average but potentially blunt - whereas adjacency-aware borrowing can target geographically localized shocks and shared dynamics among neighbors.

We propose a Bayesian multiple time-series (MTS) framework that keeps unit-level dynamics univariate (*e.g.*, $AR(1)$) while introducing cross-sectional structure via CAR priors on unit components. This setup preserves simple likelihoods, enables adjacency-aware pooling across neighbors that is crucial in data-sparse small-area settings, and delivers higher forecast accuracy with tighter uncertainty (narrower Highest Posterior Density Intervals (HPDIs), and lower dispersion). The approach remains computationally light and modular in the choice of spatial weights (adjacency) matrix, W, and it accommodates covariates and time-varying dependence to produce decision-relevant intervals for regional planning.

From a forecasting perspective, design choice matters especially when series are short, noisy, or uneven in length. Unit-wise fits may deliver wide intervals and unstable point forecasts; fully multivariate specifications can be over-parameterized and sensitive to mis-specification; and hierarchical pooling without geographic consideration can reduce variance but blur meaningful spatial gradients. The key requirement is a framework that (i) keeps per-series likelihoods simple and interpretable, (ii) borrows strength preferentially across adjacent units, (iii) scales to moderate ℓ and modest T without heavy tuning, and (iv) yields coherent predictive distributions and stable computation.

Our proposed model draws on two established strands. First, univariate dynamic modeling with hierarchical pooling for many series (West and Harrison, 1997; Prado *et al.*, 2021) supports keeping per-series likelihoods simple while sharing information across units. Second, CAR/spatio-temporal(ST)-CAR priors are standard for areal and space-time modeling, providing localized, adjacency-aware smoothing with clear conditional structure (Gelfand and Vounatsou, 2003; Jin *et al.*, 2005; Lee *et al.*, 2014; Rushworth *et al.*, 2014; Lawson, 2018; Desjardins *et al.*, 2020). Recent work strengthens this view of areal panels as collections of time series by combining CAR-based spatio-temporal effects with clustering across areas Mozdzen *et al.* (2022). See Sahu (2021) for a comprehensive treatment of Bayesian spatio–temporal models, including CAR and ST–CAR implementations in R.

In small-area estimation and panel time-series analysis, classic non-spatial pooling offers useful comparators: Fay and Herriot (1979) provide cross-sectional pooling via exchangeable area effects; Binder and Dick (1989) develop modeling and estimation for repeated surveys - effectively panels of univariate series - providing context for pooling across waves; Rao and Yu (1994) incorporate time-series evolution for area effects to borrow across time and domains; and Nandram and Petruccelli (1997) propose hierarchical Bayesian pooling for AR processes across units. For a comprehensive overview, see Rao and Molina (2015). These studies differ from our formulation by not encoding geographic adjacency - borrowing is global or exchangeable rather than spatially adaptive. When series form groups or hierarchies, forecast reconciliation ensures coherent aggregation and offers an orthogonal (non-spatial) comparators Girolimetto and Di Fonzo (2024).

Alternative routes to leveraging geography include space-time autoregressive with spatial lags (Shoesmith, 2013) and variational panel spatial autoregressive (SAR) estimators with unrestricted/learned spatial weights (Gefang *et al.*, 2025). Complementary perspectives outside the time-domain Gaussian setting include frequency-domain borrowing via spectral mixtures (Cadonna *et al.*, 2019) and nonparametric error pooling across series(Nieto-Barajas and Quintana, 2016). Bayesian models that keep series univariate yet share information hierarchically provide a non-spatial benchmark with Markov-chain Monte Carlo(MCMC) inference (Müller and Watson, 2025). Spatial Bayesian variants also refine CAR structure, including localized CAR formulations that allow smoothing to switch on/off across the map (Lee *et al.*, 2014). Recent scalable Bayesian space-time forecasting approaches (Orozco-Acosta *et al.*, 2023) provide a complementary spatial baseline focused on short-term prediction.

The proposed MTS-CAR framework does not employ likelihood-based spatial autoregressive structures such as SAR or SARAR, nor does it rely on high-dimensional VARs or black-box machine-learning forecasters. Instead, spatial dependence is introduced exclusively through a conditional autoregressive (CAR) prior on unit-specific coefficients. This enables information sharing across neighboring units while preserving simple, interpretable univariate time-series dynamics. By keeping spatial dependence in the prior rather than in the likelihood, the framework avoids the complexity and potential instability of large VAR systems or spatial lag models, while still capturing meaningful cross-sectional structure in a principled Bayesian way. As such, this study makes three key contributions. First, spatial dependence is introduced entirely through the prior distribution, rather than through the likelihood, which conceptually distinguishes the proposed framework from likelihood-based spatial autoregressive models such as SAR or SARAR. Second, spatial structure is imposed at the coefficient level, specifically on intercepts and autoregressive parameters, allowing adjacency-aware borrowing of information across neighboring units while preserving simple, interpretable univariate dynamics for each time series. Third, we adopt a collapsed Gibbs sampling strategy that is not merely a computational convenience, but a deliberate inferential design choice that reduces posterior dependence, improves mixing, and yields more efficient posterior inference.

Beyond Bayesian methods, graph neural networks learn adjacency-aware spatiotemporal dependencies and are widely used as strong forecasting baselines (Yu *et al.*, 2017), while global deep forecasters train a single model across many series and set competitive non-Bayesian baselines for long-horizon prediction (Nie *et al.*, 2023). Our MTS-CAR model remains in the time domain and encodes adjacency solely through priors on unit-level effects, yielding targeted spatial smoothing and modular computation.

Beyond proposing the model, we formalize an adjacency-aware Bayesian MTS specification with proper CAR priors on unit-level components, give clear positivity conditions for $D - \rho W$ that guarantee well-posed posteriors, and derive efficient Gibbs updates with conjugate blocks for fast computation. We also provide a clean benchmarking design against SS and non-spatial MTS baselines, report calibration diagnostics, and document robust predictive gains in data-sparse panels. The framework is modular in W and readily accommodate covariates and higher-order AR dynamics via companion-form constraints, offering a practical template for adjacency-aware forecasting in areal panels.

Across 20 areas, our MTS-CAR model delivers a median root mean square error reduction of 30.8% (maximum 50.9%) relative to the intermediate baseline model, MTS, with tighter HPDIs and lower dispersion, while preserving unit-level interpretability. Sensitivity checks indicate that these gains are robust to reasonable choices of W and prior hyperparameters.

The paper proceeds as follows: Section 2 introduces the MTS-CAR framework, detailing the likelihood and hierarchical priors including the CAR prior along with the Gibbs sampling scheme. Section 3 presents the two benchmarks: the MTS and single-series (SS) AR(1) models, along with the Bayesian estimation results and compares predictive performance across the three models. Section 4 concludes the paper with key findings, implications for spatial modeling in panel data, and avenues for future research and extensions.

2. MTS-CAR model

We introduce the proposed Bayesian multiple time series model with CAR priors (MTS-CAR). The specification retains a univariate AR(1) likelihood for each series while inducing spatial borrowing through CAR priors on the area-specific intercept and AR-slope deviations. Section 2.1 details the MTS-CAR model framework, and in Section 2.2, we describe how to run efficient Gibbs sampler with collapsed updates, which is innovative.

2.1. Model framework

For a panel of ℓ areas observed over T periods, we observe a univariate series $\{y_{it}\}_{t=1}^{T}$ and specify

$$y_{it} \;=\; \theta_0 + \nu_{0i} + \Big(\theta_1 + \nu_{1i}\Big)\, y_{i,t-1} + \varepsilon_{it}, \quad i = 1, 2, \ldots, \ell,\; t = 1, 2, \ldots, T, \tag{1}$$

where (ν_{0i}, ν_{1i}) are spatial random effects. Following Jin *et al.* (2005), with a symmetric adjacency matrix W and diagonal degree matrix D, spatial borrowing enters via CAR priors, so the CAR prior is placed on the spatial deviations ν_{0i} (intercepts) and ν_{1i} (AR coefficients).

From (1), let $\underset{\sim}{\nu}_1 = (\nu_{01}, \ldots, \nu_{0\ell})'$, and $\underset{\sim}{\nu}_2 = (\nu_{11}, \ldots, \nu_{1\ell})'$, we define the CAR prior structure $\underset{\sim}{\nu}_1 \sim \mathcal{N}(\underset{\sim}{0}, \Sigma_1)$ and $\underset{\sim}{\nu}_2 \mid \underset{\sim}{\nu}_1 \sim \mathcal{N}(A\underset{\sim}{\nu}_1, \Sigma_2)$, leading to

$$\underset{\sim}{\nu} = \begin{pmatrix} \underset{\sim}{\nu}_1 \\ \underset{\sim}{\nu}_2 \end{pmatrix} \sim \mathcal{N}(\underset{\sim}{0}, \boldsymbol{\Sigma}),$$

with the covariance matrix and the precision matrix define as

$$\mathbf{\Sigma} = \begin{pmatrix} \Sigma_1 & \Sigma_1 A' \\ A\Sigma_1 & \Sigma_2 + A\Sigma_1 A' \end{pmatrix}, \qquad \mathbf{\Sigma}^{-1} = \begin{pmatrix} A'\Sigma_2^{-1}A + \Sigma_1^{-1} & -A'\Sigma_2^{-1} \\ -\Sigma_2^{-1}A & \Sigma_2^{-1} \end{pmatrix},$$

where $\Sigma_s = \tau_s^2(D - \rho_s W)^{-1}$ for $s = 1, 2$. D is a diagonal matrix containing the row/column sums of W as diagonal elements. For objective-Bayes guidance on propriety and prior specification with intrinsic condition autoregressive (ICAR) components, see Keefe *et al.* (2019). The matrix $A = (m_0 I + m_1 W)$ and $\frac{1}{\lambda_{(1)}} \leq \rho_s \leq \frac{1}{\lambda_{(\ell)}}$, such that $\lambda_{(1)} < \cdots < \lambda_{(\ell)}$ are eigenvalues of $D^{-1}W$. Under standard spectral bounds for ρ_s, $D - \rho_s W$ is positive definite, yielding a proper Gaussian prior and stable computations. We show how to obtain the joint distribution of $\underset{\sim}{\nu} = (\underset{\sim}{\nu}_1,\ \underset{\sim}{\nu}_2)'$ in Appendix A (supplement), and the derivation of the covariance matrix in Appendix B (supplement). See Gelfand and Vounatsou (2003) and Jin *et al.* (2005) for more details on multivariate CAR models. We assign flat or non-informative priors to the hyperparameters and latent initial condition, y_{i0} as

$$\pi(\theta_0, \theta_1, \sigma^2, m_0, m_1, \tau_1^2, \tau_2^2, \rho_1, \rho_2) \propto \frac{1}{\sigma^2} \cdot \frac{1}{\tau_1^2} \cdot \frac{1}{\tau_2^2}, \qquad y_{i0} \sim \mathcal{N}(\mu_0, \sigma_0^2),$$

where μ_0, the prior mean and σ_0^2 the prior variance of y_{0i}, are to be specified.

For simplicity, we define $X = (1,\ y_{i,t-1})'$, $\underset{\sim}{\theta} = (\theta_0,\ \theta_1)'$; $Z = (Z_0,\ Z_1) \in \mathbb{R}^{\ell T \times 2\ell}$, where $Z_0 \in \mathbb{R}^{\ell T \times \ell}$ is an indicator matrix that maps $\underset{\sim}{\nu}_1$ to the appropriate rows (1 if the row belongs to unit i, otherwise 0), and $Z_1 \in \mathbb{R}^{\ell T \times \ell}$ has the same pattern as Z_0 but with entries $y_{i,t-1}$ in column i for rows of unit i, mapping $\underset{\sim}{\nu}_2$ to the appropriate row. Let $\underset{\sim}{y} = (y_{1t}, \ldots, y_{\ell t})' \in \mathbb{R}^{T \times \ell}$ (stacked over t). Then (1) becomes $\underset{\sim}{y} = X\underset{\sim}{\theta} + Z\underset{\sim}{\nu} + \underset{\sim}{\varepsilon}$. By Bayes' theorem, we obtained the joint posterior distribution as

$$\begin{aligned} \pi\Big(\underset{\sim}{\nu}, \underset{\sim}{\theta}, \sigma^2, &m_0, m_1, \tau_1^2, \tau_2^2, \rho_1, \rho_2, \underset{\sim}{y}_0 \mid \underset{\sim}{y}\Big) \propto \pi(\underset{\sim}{y}_0) \\ &\times \left(\frac{1}{\sigma^2}\right)^{\frac{T\ell}{2}+1} \exp\Big\{-\frac{1}{2\sigma^2}(\underset{\sim}{y} - X\underset{\sim}{\theta} - Z\underset{\sim}{\nu})'(\underset{\sim}{y} - X\underset{\sim}{\theta} - Z\underset{\sim}{\nu})\Big\} \\ &\times (\frac{1}{\tau_1^2})^{\frac{\ell}{2}+1}|(D - \rho_1 W)|^{1/2}\ (\frac{1}{\tau_2^2})^{\frac{\ell}{2}+1}|(D - \rho_2 W)|^{1/2}\ \exp\Big\{-\tfrac{1}{2}\underset{\sim}{\nu}'\mathbf{\Sigma^{-1}}\underset{\sim}{\nu}\Big\}. \end{aligned} \tag{2}$$

We show the computations of $\det(\mathbf{\Sigma})$ and Σ_s^{-1} in Appendix C and D (supplement), respectively.

2.2. Conditional posteriors and Gibbs updates

Now, we show how to fit the posterior density described in (2) using the Gibbs sampler. We derive the full conditional posterior distributions (CPDs) required for the Gibbs sampler. Let the unknowns parameters be partitioned as $\{\underset{\sim}{\theta}, \underset{\sim}{\nu}, \sigma^2, \tau_1^2, \tau_2^2, \rho_1, \rho_2, \underset{\sim}{y}_0\}$. For each block we derive the conditional kernel, up to a proportionality constant, with respect to the joint posterior, and identify the sampling family when conjugacy holds. Parameters without closed-form conditionals (*e.g.*, ρ_1, ρ_2) are updated by a fixed grid step.

The initial condition y_{i0} affects the first time step in the autoregressive process. Its CPD is derived from both the likelihood and its normal prior. From (2), when $t = 1$, the observation equation in area i reads

$$y_{i1} = \theta_0 + \nu_{0i} + (\theta_1 + \nu_{1i}) y_{i0} + \varepsilon_{i1}.$$

With a weakly informative Gaussian prior, $\pi(y_{i0})$, the conditional posterior, a Gaussian, is obtained by completing the square as

$$y_{i0} \mid \theta_0, \theta_1, \nu_{0i}, \nu_{1i}, \sigma^2, y_{i1} \sim \mathcal{N}\Big(A^{-1}B,\ A^{-1}\Big), \tag{3}$$

where

$$A = \frac{1}{\sigma^2}(\theta_1 + \nu_{1i})^2 + \frac{1}{\sigma_0^2}; \qquad B = \frac{1}{\sigma^2}(\theta_1 + \nu_{1i})\Big[y_{i1} - (\theta_0 + \nu_{0i})\Big] + \frac{\mu_0}{\sigma_0^2}.$$

The posterior mean is a precision-weighted compromise between (i) the "reverse regression" estimate implied by the $AR(1)$ measurement and (ii) the prior mean μ_0. When the local persistence is large or the observation noise σ^2 is small, the likelihood contribution dominates; otherwise the prior stabilizes y_{0i}. This update makes the first time row coherent with the area-specific dynamics before the latent spatial effects are drawn.

The observation variance, σ^2 is the model's noise level, measuring idiosyncratic dispersion around each series' conditional mean and governing forecast uncertainty. From (2), we have

$$\pi(\sigma^2 \mid \underset{\sim}{\theta}, \underset{\sim}{\nu}, m_0, m_1, \tau_1^2, \tau_2^2, \rho_1, \rho_2, \underset{\sim}{y}_0, \underset{\sim}{y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T\ell}{2}+1} \exp\Big\{-\frac{1}{2\sigma^2}(\underset{\sim}{y} - X\underset{\sim}{\theta} - Z\underset{\sim}{\nu})'(\underset{\sim}{y} - X\underset{\sim}{\theta} - Z\underset{\sim}{\nu})\Big\}.$$

Let $K = (\underset{\sim}{y} - X\underset{\sim}{\theta} - Z\underset{\sim}{\nu})'(\underset{\sim}{y} - X\underset{\sim}{\theta} - Z\underset{\sim}{\nu})$. Then

$$\sigma^2 \mid \underset{\sim}{\theta}, \underset{\sim}{\nu}, m_0, m_1, \tau_1^2, \tau_2^2, \rho_1, \rho_2, \underset{\sim}{y}_0, \underset{\sim}{y} \sim \mathrm{IG}\left(\frac{T\ell}{2}, \frac{K}{2}\right). \tag{4}$$

We update the common coefficients $\underset{\sim}{\theta}$ and the area-specific effects $\underset{\sim}{\nu}$ in a single block using a collapsed-then-conditional scheme. From (2),we have

$$\pi\Big(\underset{\sim}{\nu}, \underset{\sim}{\theta} \mid \underset{\sim}{y}\Big) \propto \exp\Big\{-\frac{1}{2\sigma^2}(\underset{\sim}{y} - X\underset{\sim}{\theta} - Z\underset{\sim}{\nu})'(\underset{\sim}{y} - X\underset{\sim}{\theta} - Z\underset{\sim}{\nu})\Big\} \exp\Big\{-\tfrac{1}{2}\underset{\sim}{\nu}'\boldsymbol{\Sigma^{-1}}\underset{\sim}{\nu}\Big\}.$$

We sampled $(\underset{\sim}{\nu},\ \underset{\sim}{\theta})$ using $\pi(\underset{\sim}{\nu},\ \underset{\sim}{\theta}) = \pi(\underset{\sim}{\nu} \mid \underset{\sim}{\theta}, \sigma^2, \underset{\sim}{y})\pi(\underset{\sim}{\theta} \mid \sigma^2, \underset{\sim}{y})$. To do this, we marginalize $\underset{\sim}{\nu}$ and defined $V = \sigma^2 I_n + Z\boldsymbol{\Sigma}Z'$, the conditional posterior is Gaussian

$$\underset{\sim}{\theta} \mid \underset{\sim}{y}, \sigma^2 \sim \mathcal{N}\Big(\mu_\theta, \Sigma_\theta\Big), \quad \text{where} \quad \Sigma_\theta = \Big(X'V^{-1}X\Big)^{-1}, \quad \mu_\theta = \Sigma_\theta X'V^{-1}y. \tag{5}$$

Conditioning on the newly sampled $\underset{\sim}{\theta}$, the effects $\underset{\sim}{\nu}$ are again Gaussian

$$\underset{\sim}{\nu} \mid \underset{\sim}{\theta}, \sigma^2, \underset{\sim}{y} \sim \mathcal{N}\Big(V_0^{-1}V_a,\ V_0^{-1}\Big), \text{ where } V_0 = \Big[\tfrac{1}{\sigma^2}Z'Z + \boldsymbol{\Sigma^{-1}}\Big],\ V_a = \Big[\tfrac{1}{\sigma^2}Z'(\underset{\sim}{y} - X\underset{\sim}{\theta})\Big]. \tag{6}$$

Sampling $\underset{\sim}{\theta}$ after integrating out the high-dimensional $\underset{\sim}{\nu}$ removes the strong posterior dependence between $\underset{\sim}{\theta}$ and $\underset{\sim}{\nu}$ that otherwise causes high auto-correlation and slow exploration. The collapsed likelihood has well-behaved curvature (through V), so $\underset{\sim}{\theta}$ updates adapt to the uncertainty already induced by the spatial effects. We then condition on $\underset{\sim}{\theta}$ and draw $\underset{\sim}{\nu}$ from its conjugate Gaussian CPD. Empirically this collapsed-then-conditional order substantially reduces within-chain auto-correlation for $\underset{\sim}{\theta}$ while retaining simple, stable linear-algebra for $\underset{\sim}{\nu}$. Within each MCMC iteration we jointly treat $(\underset{\sim}{\theta},\ \underset{\sim}{\nu})$ as one block updated in two substeps: (i) sample $\underset{\sim}{\theta}$ from (5) above; (ii) immediately sample $\underset{\sim}{\nu}$ given this $\underset{\sim}{\theta}$. No other parameters are updated in between these two substeps.

The hyperparameters, τ_s^2 sets the overall precision of the latent spatial effects. From (2), we have the terms with τ_s^2 as

$$\pi(\tau_s^2 \mid \theta, \underset{\sim}{\nu}, m_0, m_1, \rho_1, \rho_2, \underset{\sim}{y}_0, \underset{\sim}{y}) \ \propto \ (\frac{1}{\tau_s^2})^{\ell/2+1} \ \exp\Big\{-\tfrac{1}{2}\underset{\sim}{\nu}'\mathbf{\Sigma^{-1}}\underset{\sim}{\nu}\Big\}\,,$$

expanding $\underset{\sim}{\nu}'\mathbf{\Sigma^{-1}}\underset{\sim}{\nu}$ with $\Sigma_s^{-1} = \frac{1}{\tau_s^2}(D-\rho_s W)$, and after a little algebra we have

$$\pi(\tau_s^2 \mid \underset{\sim}{\theta}, \underset{\sim}{\nu}, m_0, m_1, \rho_1, \rho_2, \underset{\sim}{y}_0, \underset{\sim}{y}) \ \propto \ (\frac{1}{\tau_s^2})^{\ell/2+1} \ \exp{-\tfrac{1}{2}}\left\{\frac{1}{\tau_s^2}Q_s\right\}.$$

Thus,

$$\tau_s^2 \mid \underset{\sim}{\theta}, \underset{\sim}{\nu}, m_0, m_1, \rho_1, \rho_2, \underset{\sim}{y}_0, \underset{\sim}{y} \sim \text{Inverse} - \text{Gamma}\left(\frac{\ell}{2},\, \frac{1}{2}Q_s\right), \qquad s = 1, 2. \tag{7}$$

where

$$Q_1 = \underset{\sim}{\nu}_1'(D-\rho_1 W)\underset{\sim}{\nu}_1, \qquad Q_2 = (\underset{\sim}{\nu}_2 - A\underset{\sim}{\nu}_1)'(D-\rho_2 W)(\underset{\sim}{\nu}_2 - A\underset{\sim}{\nu}_1).$$

Because $\Sigma_s = \tau_s^2(D-\rho_s W)^{-1}$, the covariance increases linearly with τ_s^2. Consequently, larger τ_s^2 reduces precision and weakens smoothing (less pooling across areas, more spatial variability), whereas smaller τ_s^2 increases precision and strengthens smoothing (more pooling, smoother latent fields). Under the inverse-gamma conditional posterior for τ_s^2 with a scale proportional to the quadratic form $a_s'(D-\rho_s W)a_s$, larger quadratic forms shift mass toward larger τ_s^2 (weaker smoothing), while smaller quadratic forms shift mass toward smaller τ_s^2 (stronger smoothing).

With an admissible interval $\Big(1/\lambda_{(1)},\ 1/\lambda_{(\ell)}\Big)$, where $\lambda_{(1)}$ and $\lambda_{(\ell)}$ are the minimum and maximum eigenvalues of $D^{-1}W$, respectively, ρ_s governs the strength and effective range of spatial coupling defined by the adjacency structure. From (2), and using the expansion of $\underset{\sim}{\nu}'\mathbf{\Sigma}^{-1}\underset{\sim}{\nu}$, we have

$$\begin{aligned}
&\pi(\rho_s \mid \underset{\sim}{\theta}, \underset{\sim}{\nu}, \sigma^2, m_0, m_1, \tau_s^2, \underset{\sim}{y}_0, \underset{\sim}{y}) \propto \\
&\qquad |D-\rho_s W|^{1/2} \ \exp{-\tfrac{1}{2}}\left\{\frac{1}{\tau_s^2}\underset{\sim}{a}_s'(D-\rho_s W)\underset{\sim}{a}_s\right\}, \\
&\qquad \propto \left(\prod_{i=1}^{\ell} d_{ii}\right)^{1/2} \prod_{k=1}^{\ell}\Big(1-\rho_s\lambda_k\Big)^{1/2} \ \exp{-\tfrac{1}{2}}\left\{\frac{1}{\tau_s^2}\underset{\sim}{a}_s'(D-\rho_s W)\underset{\sim}{a}_s\right\},
\end{aligned} \tag{8}$$

where $\underset{\sim}{a}_1 = \underset{\sim}{\nu}_1$, $\underset{\sim}{a}_2 = \underset{\sim}{\nu}_2 - A\underset{\sim}{\nu}_1$, and λ_k is an eigenvalues of $D^{-1}W$. In the marginal likelihood, the log-determinant term penalizes values outside the admissible, stable region, while the quadratic form rewards values that reduce spatial roughness in a_s. Small ρ_s implies weak dependence and minimal smoothing across neighbors; larger ρ_s induces stronger, longer-range smoothing over the graph, subject to the positive-definiteness constraints determined by the adjacency matrix.

For the coupling parameters $(m_0,\ m_1)$, the only dependence on m_0, m_1 is through $A = m_0 I + m_1 W$ inside $\mathbf{\Sigma^{-1}}$. Let $\underset{\sim}{m} = (m_0,\ m_1)'$, from (2), we have

$$\pi(\underset{\sim}{m} \mid \underset{\sim}{\theta}, \underset{\sim}{\nu}, \sigma^2, \rho_s, \tau_s^2, \underset{\sim}{y}_0, \underset{\sim}{y}) \propto \exp\Big\{-\tfrac{1}{2}\underset{\sim}{\nu}'\mathbf{\Sigma^{-1}}\underset{\sim}{\nu}\Big\}\,.$$

In the expansion of $\underset{\sim}{\nu}'\mathbf{\Sigma^{-1}}\underset{\sim}{\nu}$, the only term having m_0 and m_1 through A are $(\underset{\sim}{\nu}_2 - A\underset{\sim}{\nu}_1)'\Sigma_2^{-1}(\underset{\sim}{\nu}_2 - A\underset{\sim}{\nu}_1)$. Expanding this term we have $\underset{\sim}{\nu}_2'\Sigma_2^{-1}\underset{\sim}{\nu}_2 - 2\,\underset{\sim}{\nu}_1'A'\Sigma_2^{-1}\underset{\sim}{\nu}_2 + \underset{\sim}{\nu}_1'A\Sigma_2^{-1}A'\underset{\sim}{\nu}_1$. Let $U = m_0 I$ and $B = m_1 W$. Then $\underset{\sim}{\nu}_1'A\Sigma_2^{-1}A'\underset{\sim}{\nu}_1 = \underset{\sim}{\nu}_1'\Big(U'\Sigma_2^{-1}U + U'\Sigma_2^{-1}B + B'\Sigma_2^{-1}U + B'\Sigma_2^{-1}B\Big)\underset{\sim}{\nu}_1$. Therefore

$$\begin{aligned}(\underset{\sim}{\nu}_2 - A'\underset{\sim}{\nu}_1)'\Sigma_2^{-1}(\underset{\sim}{\nu}_2 - A'\underset{\sim}{\nu}_1) &= \underset{\sim}{\nu}_2'\Sigma_2^{-1}\underset{\sim}{\nu}_2 \;-\; 2\,\underset{\sim}{\nu}_1'U'\Sigma_2^{-1}\underset{\sim}{\nu}_2 \;-\; 2\,\underset{\sim}{\nu}_1'B'\Sigma_2^{-1}\underset{\sim}{\nu}_2 \\ &\quad + \;\underset{\sim}{\nu}_1'U\Sigma_2^{-1}U'\underset{\sim}{\nu}_1 \;+\; \underset{\sim}{\nu}_1'U'\Sigma_2^{-1}B\underset{\sim}{\nu}_1 \;+\; \underset{\sim}{\nu}_1'B'\Sigma_2^{-1}U\underset{\sim}{\nu}_1 \;+\; \underset{\sim}{\nu}_1'B'\Sigma_2^{-1}B\underset{\sim}{\nu}_1.\end{aligned}$$

Substituting $U = m_0 I$ and $B = m_1 W$, and pulling out scalars m_0, m_1, we have the whole expression as the quadratic polynomial in $(m_0,\ m_1)$ as

$$Q(m_0, m_1) = m_0^2\, a + m_1^2\, c + 2m_0 m_1\, b - 2m_0\, d - 2m_1\, e + C,$$

where $a = \underset{\sim}{\nu}_1'\,\Sigma_2^{-1}\,\underset{\sim}{\nu}_1,\ b = \underset{\sim}{\nu}_1'\,\Sigma_2^{-1}W\,\underset{\sim}{\nu}_1,\ c = \underset{\sim}{\nu}_1'\,W\Sigma_2^{-1}W\,\underset{\sim}{\nu}_1,\ d = \underset{\sim}{\nu}_1'\,\Sigma_2^{-1}\,\underset{\sim}{\nu}_2, e = \underset{\sim}{\nu}_1'\,W\Sigma_2^{-1}\,\underset{\sim}{\nu}_2,\ C = \underset{\sim}{\nu}_2'\Sigma_2^{-1}\underset{\sim}{\nu}_2, W' = W$. Let $\underset{\sim}{h} = \begin{pmatrix} d \\ e \end{pmatrix}$, and $G = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, then $Q(m_0, m_1) \;=\; \underset{\sim}{m}'G\underset{\sim}{m} \;-\; 2\underset{\sim}{h}'\underset{\sim}{m} \;+\; C$. Thus we have

$$\pi(\underset{\sim}{m} \mid \underset{\sim}{\theta}, \underset{\sim}{\nu}, \sigma^2, \rho_s, \tau_s^2, \underset{\sim}{y}_0, \underset{\sim}{y}) \propto \exp\Big\{-\tfrac{1}{2}\,Q(m_0, m_1)\Big\},$$

which has the bivariate normal kernel $\exp\Big\{-\tfrac{1}{2}\,(\underset{\sim}{m} - \mu)'G(\underset{\sim}{m} - \mu)\Big\}$ with $\mu = G^{-1}\underset{\sim}{h}$ and $\mathrm{Cov}(\underset{\sim}{m}) = G^{-1}$. Thus

$$\underset{\sim}{m} \mid \underset{\sim}{\theta}, \underset{\sim}{\nu}, \sigma^2, \rho_s, \tau_s^2, \underset{\sim}{y}_0, \underset{\sim}{y} \;\sim\; \mathcal{N}\Big(G^{-1}\underset{\sim}{h},\ G^{-1}\Big), \tag{9}$$

where $\widetilde{\Delta} = \Big(\underset{\sim}{\nu}_1'\widetilde{\Sigma}\underset{\sim}{\nu}_2\Big)\Big(\underset{\sim}{\nu}_2'W\widetilde{\Sigma}W\underset{\sim}{\nu}_2\Big) - \Big(\underset{\sim}{\nu}_2'\widetilde{\Sigma}W\underset{\sim}{\nu}_2\Big)^2, \qquad G^{-1} = \frac{1}{\widetilde{\Delta}}\begin{pmatrix} \underset{\sim}{\nu}_2'W\widetilde{\Sigma}W\underset{\sim}{\nu}_2 & -\underset{\sim}{\nu}_2'\widetilde{\Sigma}W\underset{\sim}{\nu}_2 \\ -\underset{\sim}{\nu}_2'\widetilde{\Sigma}W\underset{\sim}{\nu}_2 & \underset{\sim}{\nu}_2'\widetilde{\Sigma}\underset{\sim}{\nu}_2 \end{pmatrix},$

$$G^{-1}\underset{\sim}{h} = \frac{1}{\widetilde{\Delta}}\begin{pmatrix} \Big(\underset{\sim}{\nu}_2'W\widetilde{\Sigma}W\underset{\sim}{\nu}_2\Big)\Big(\underset{\sim}{\nu}_2'\widetilde{\Sigma}\underset{\sim}{\nu}_1\Big) - \Big(\underset{\sim}{\nu}_2'\widetilde{\Sigma}W\underset{\sim}{\nu}_2\Big)\Big(\underset{\sim}{\nu}_2'W\widetilde{\Sigma}\underset{\sim}{\nu}_1\Big) \\ \Big(\underset{\sim}{\nu}_2'\widetilde{\Sigma}\underset{\sim}{\nu}_2\Big)\Big(\underset{\sim}{\nu}_2'W\widetilde{\Sigma}\underset{\sim}{\nu}_1\Big) - \Big(\underset{\sim}{\nu}_2'\widetilde{\Sigma}W\underset{\sim}{\nu}_2\Big)\Big(\underset{\sim}{\nu}_2'\widetilde{\Sigma}\underset{\sim}{\nu}_1\Big) \end{pmatrix}, \qquad \widetilde{\Sigma} := (D - \rho_2 W)^{-1}.$$

Under the squared-scale precision parameterization, the common multiplier τ_2^2 cancels from this conditional; hence the mean and precision above are invariant to the change $\Sigma_2^{-1} = \frac{1}{\tau_2^2}(D - \rho_2 W)$. We show the derivation of G^{-1} and $G^{-1}\underset{\sim}{h}$ in Appendix E (supplement). For the joint distribution of $\underset{\sim}{m}$ to be well-defined (*i.e.*, G invertible/PD), we need $a > 0$, $c > 0$, $\det(G) = \widetilde{\Delta} = ac - b^2 > 0$, which are algebraic conditions on the scalars a, b, c. We prove that these conditions are satisfied almost surely in Appendix F (supplement).

Finally, we made a few numerical remarks. The matrix $D - \rho W$ is positive definite for $\rho \in \left(1/\lambda_{(1)},\, 1/\lambda_{(\ell)}\right)$ (proved in Appendix G (supplement)) and $D - I$ is positive semi-definite, *i.e.* $D \succeq I$, then the CAR precision components are symmetric positive definite (SPD); consequently, the block prior precision $\mathbf{\Sigma^{-1}}$ is SPD as well. We operate only in this admissible region, so the joint prior for the latent effects is well-behaved. In the collapsed updates, we never form large covariance matrices explicitly. instead we apply the Sherman-Morrison-Woodbury identity (Sherman and Morrison, 1950; Woodbury, 1950; Henderson and Searle, 1981; Golub and Van Loan, 2013) to reduce computations to solving linear systems with a single symmetric positive-definite matrix and performing a small 2×2 Cholesky decomposition for the regression block. This keeps the algebra stable, avoids explicit matrix inverses, and yields a fast, numerically reliable implementation.

3. Data analysis

We assess the MTS-CAR model's performance against the two baselines using a panel of annual average hourly earnings for production employees across $\ell = 20$ California Metropolitan Statistical Areas (MSAs), spanning 2003–2024 (22 years). To encode geography, we build a spatial incidence (adjacency) matrix capturing proximity and inter-regional links among MSAs. Because autoregressive specifications requires stationarity, we applied a square-root transformation to enforce positivity and then performed second-difference of the transformed series (after performing unit root/stationarity test using Augmented Dickey-Fuller (ADF) and Phillips-Perron tests). We do not use the logarithmic transformation because when the logarithm is re-transformed, the Bayesian predictive distribution is not well defined; *e.g.*, see Manandhar and Nandram (2021, 2025) who used generalized gamma distributions instead of the logarithmic transformation. For evaluation, we reserve the final time point in each series as a one-step-ahead holdout and use this left-out observation's forecast to compare predictive performance across models.

For the panel analysis, the SS model relied on a simple random sample of 1,000 posterior draws. For the MTS model, we ran 75,000 Gibbs iterations, discarded the first 15,000 as burn-in, and thinned by keeping every 60th draw to obtain 1,000 samples. For the MTS-CAR model, we ran 13,000 iterations, dropped the first 3,000, and kept every 10th draw - again yielding 1,000 posterior samples to keep sample sizes comparable across models. The MTS-CAR Gibbs sampler was executed in R-studio on a Windows 11 desktop computer equipped with Intel(R) Core(TM) i7-10700 CPU 2.90GHz processor and 16 GB RAM. A full MCMC run of 13,000 iterations took approximately 48 minutes.

After sampling, we assessed convergence using the Geweke and Gelman-Rubin diagnostic tests for each parameter. We also computed the effective sample size (ESS) to quantify the amount of independent information and examined trace plots to visually evaluate mixing of the Markov chain. For the MTS Gibbs samplers, Geweke p-values ranged from 0.074 to 0.999, and the smallest ESS across parameters was 855, indicating strong convergence. Table 1 presents the ESS value, the Geweke test, and the Gelman-Rubin test of convergence for the hyperparameters in the MTS-CAR model. As shown in the table, the ESS (> 950) for all the hyperparameters are large, and the Geweke test p-value (> 0.05) indicates that these

Table 1: Convergence diagnostics

Parameter	ESS	Geweke-test	PSRF
θ_0	1000	0.814	1.001
θ_1	1694	0.954	1.000
m_0	1100	0.959	0.999
m_1	1064	0.700	1.009
ρ_1	1000	0.377	1.001
ρ_2	959	0.810	1.001
τ_1^2	1000	0.083	1.000
τ_2^2	1000	0.612	1.000
σ^2	984	0.722	0.999

Note: ESS = Effective Sample Size; PSRF = Potential Scale Reduction Factor; the Geweke test p-value was reported.

parameters converges well. The p-value of the Geweke test of the remaining parameters and latent variable (e.g, $\nu_{0i}, \nu_{1i}, y_{0i}$) ranged from 0.111 to 0.985, with the smallest ESS value of 902. As reported in Table 1, the potential scale reduction factor ($PSRF < 1.01$) for all the hyperparameters have excellent convergence. The remaining parameters show excellence convergence ($PSRF \leq 1.01$) as well.

3.1. Predictive evaluation and forecasting accuracy

To evaluate the predictive performance of the MTS-CAR model, we made prediction of the last observation ($t = T$) for each series. These predictions were computed based on posterior draws from the Gibbs sampler. Specifically, using $\hat{y}_{iT|T-1} = \theta_0 + \nu_{0i} + \Big(\theta_1 + \nu_{1i}\Big)\, y_{i,T-1} + \varepsilon_{iT},$, we predicted the last observation for each series $i = 1, \ldots, \ell$. We also made predictions using the baseline SS and MTS models.

Since our estimation used square-root–transformed, and second-differenced data, we mapped forecasts back to the original scale via the inverse transformations. Specifically, we first undid the second differencing using the past values $\sqrt{y_{T-1}}$ and $\sqrt{y_{T-2}}$ to recover $\sqrt{y_T}$, and then squared the result to invert the square-root transform. We summarized posterior predictions using the posterior mean (PM), posterior standard deviation (PSD), posterior bias (PB), posterior root mean squared error (RMSE), posterior coefficient of variation (PCV), and highest posterior density interval (HPDI), providing a comprehensive assessment of each forecasting approach.

In addition to posterior predictive summaries, we compute prediction relative errors as

$$R_{iT} \; = \; \frac{\hat{y}_{iT} - y_{iT}}{y_{iT}},$$

where $\hat{y}_{iT}$ is the predicted level and y_{iT} is the observed value at the final time point. From these errors, we report summary measures including the posterior error mean (PEM), posterior error standard deviation (PESD), Bayesian standard deviation (BSTD), and the 95% HPDI. These metrics are presented separately for the SS model, the MTS model without spatial coupling , and the proposed MTS-CAR model.

3.2. Model comparisons

Here we introduce the two baseline models used to contextualize the performance of the MTS-CAR model. The first is the SS $AR(1)$ model fitted independently to each unit/MSA, which preserves per-series interpretability but ignores cross-unit structure. The second is the MTS model that pools information across units through exchangeable hierarchical priors on intercepts and AR coefficients, yet imposes no spatial prior. Both baselines generate posterior predictive distributions against which we compare MTS-CAR model in terms of accuracy and calibration.

3.2.1. Single-Series (SS) model

As a baseline, we use a SS AR(1) for a univariate sequence $\{y_t\}_{t=1}^{T}$. The specification is

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \quad t = 1, 2, \ldots, T.$$

To simplify estimation, the model was described in matrix form as

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}, \quad \underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma^2 \sim \mathcal{N}(X\underset{\sim}{\beta}, \sigma^2 I_T),$$

where X is the $T \times 2$ design matrix with a column of ones (intercept) and the lagged values y_{t-1}, $\underset{\sim}{y} = (y_1, \ldots, y_T)'$, $\underset{\sim}{\beta} = (\beta_0, \beta_1)'$, and I_T denotes the $T \times T$ identity matrix. To complete the Bayesian setup, we assign the following prior distributions to the parameters:

$$\underset{\sim}{\beta} \mid \sigma^2 \sim \mathcal{N}(\boldsymbol{\beta}_0, \sigma^2 \Sigma_0), \quad \sigma^2 \sim \text{Inverse-Gamma}\left(\frac{a_0}{2}, \frac{b_0}{2}\right).$$

where $\boldsymbol{\beta}_0$ is the prior mean of $\underset{\sim}{\beta}$, Σ_0 is its prior scale/covariance, and a_0 and b_0 are the shape and scale, respectively, for the inverse-gamma prior on σ^2. These choices are the standard conjugate priors for the regression coefficients and the error variance, enabling straightforward Gibbs sampling. The hyperparameters a_0, and b_0 are fixed a priori, while $\boldsymbol{\beta}_0$ and Σ_0 are obtained from ordinary least squares estimate and held fixed during posterior sampling. Because y_0 is required to start the recursion in X, we place the following prior on y_0:

$$y_0 \sim \mathcal{U}(c_0, d_0), \quad \text{where } c_0 = 0.75 \min(y_t), \quad d_0 = 1.25 \max(y_t).$$

This uniform prior over a scaled range of the observed series offers weakly informative yet plausible support for the initial value y_0. Applying Bayes' theorem and collecting exponential terms involving $\underset{\sim}{\beta}$, the joint posterior simplifies to

$$\pi(\underset{\sim}{\beta}, \sigma^2, y_0 \mid \underset{\sim}{y}) \propto \left(\tfrac{1}{\sigma^2}\right)^{\left(\frac{T+a_0+2}{2}+1\right)}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{\beta} - \underset{\sim}{\hat{\beta}}_*)'(X'X + \Sigma_0^{-1})(\underset{\sim}{\beta} - \underset{\sim}{\hat{\beta}}_*)\right\} \times \exp\left\{-\frac{1}{2\sigma^2} b_{(y_0)}\right\}, \tag{10}$$

where

$$b_{(y_0)} = (\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})'(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}) + (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}_0)'X'X(X'X + \Sigma_0^{-1})^{-1}\Sigma_0^{-1}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}_0) + b_0.$$

We describe the estimation of (10) via a random sampler in Appendix H (supplement).

3.2.2. Multiple time series (MTS) model

We now consider a multiple time series (MTS) model, where a separate AR(1) process is fitted for each area $i = 1, \ldots, \ell$, while introducing hierarchical priors that share hyperparameters across units, thereby borrowing strength through the prior distribution. The model accounts for heterogeneity in intercepts and autoregressive slopes across units by placing hierarchical priors on the parameters. Specifically, let $i = 1, \ldots, \ell$ index area and $t = 1, \ldots, T$ index time. The observation model is a series-specific AR(1) regression with random intercept, μ_i and slope, b_i,

$$y_{it} \mid \mu_i, b_i, \sigma^2, y_{i,t-1} \sim \mathcal{N}\Big(\mu_i + b_i y_{i,t-1},\ \sigma^2\Big), \qquad t = 1, \ldots, T, \quad i = 1, \ldots, \ell.$$

$$\begin{aligned} \mu_i &\sim \mathcal{N}(\theta_1, \sigma_\mu^2), \\ b_i \mid \mu_i &\sim \mathcal{N}\left(\theta_2 + \rho\frac{\sigma_b}{\sigma_\mu}(\mu_i - \theta_1),\ (1-\rho^2)\sigma_b^2\right). \end{aligned}$$

where θ_1 and θ_2 are global mean parameters, $\sigma_\mu^2 > 0$ and $\sigma_b^2 > 0$ are variance hyperparameters, and $\rho \in (-1, 1)$ models correlation between μ_i and b_i. The factor σ_b/σ_μ ensures that ρ is indeed the correlation between μ_i and b_i. Let $\underset{\sim}{\beta}_i = (\mu_i, b_i)$ and $\underset{\sim}{m}_0 = (\theta_1, \theta_2)'$. The random effects follow a bivariate Normal distribution

$$\underset{\sim}{\beta}_i \mid \underset{\sim}{m}_0, \sigma_\mu^2, \sigma_b^2, \rho \sim \mathcal{N}_2\Big(\underset{\sim}{m}_0, \hat{\Sigma}\Big), \qquad \hat{\Sigma} = \begin{pmatrix} \sigma_\mu^2 & \rho\sigma_\mu\sigma_b \\ \rho\sigma_\mu\sigma_b & \sigma_b^2 \end{pmatrix},$$

The initial latent condition is unknown, so we assign a prior:

$$y_{i0} \sim \mathcal{N}(\mu_0, \sigma_0^2), \qquad \mu_0, \sigma_0^2 \text{ are obtained from a pre-sample.}$$

Weakly informative or diffuse priors are assumed for the remaining hyperparameters:

$$\pi(\theta_1, \theta_2) \propto 1, \quad \pi(\sigma^2) \propto \frac{1}{\sigma^2}, \quad \pi(\sigma_\mu^2) \propto \frac{1}{(1+\sigma_\mu^2)^2}, \quad \pi(\sigma_b^2) \propto \frac{1}{(1+\sigma_b^2)^2}, \quad \rho \sim \text{Unif}(-1, 1).$$

To enable grid sampling on a compact domain, $(0, 1)$, we define and induced marginal priors for τ_μ, τ_b as

$$\tau_\mu := \frac{1}{1+\sigma_\mu^2}, \qquad \tau_b := \frac{1}{1+\sigma_b^2}, \qquad \tau_\mu, \tau_b \in (0, 1).$$

The inverse mapping is

$$\sigma_\mu^2 = \frac{1-\tau_\mu}{\tau_\mu}, \qquad \sigma_b^2 = \frac{1-\tau_b}{\tau_b}, \qquad |J| = \left|\frac{d\sigma^2}{d\tau}\right| = \frac{1}{\tau^2}.$$

Hence,

$$\pi(\tau) \propto \pi(\sigma^2(\tau)) \left| \frac{d\sigma^2}{d\tau} \right| \propto \left(\frac{1}{1+\sigma^2(\tau)} \right)^2 \cdot \frac{1}{\tau^2} = \tau^2 \cdot \frac{1}{\tau^2} = 1,$$

so $\tau_\mu \sim \text{Unif}(0,1), \tau_b \sim \text{Unif}(0,1)$. To write the model in matrix form, we define the $T \times 2$ design matrix $X_i = [\mathbf{1}_T, \mathbf{y}_{i,-T}]$ and write the area-specific observation vector as $\underset{\sim}{y}_i = (y_{i1}, \ldots, \ y_{iT})'$. Then, the likelihood model can be expressed compactly as

$$\underset{\sim}{y}_i = X_i \underset{\sim}{\beta}_i + \underset{\sim}{\varepsilon}_i, \quad \underset{\sim}{\varepsilon}_i \sim \mathcal{N}_T(\underset{\sim}{0}, \sigma^2 \mathbf{I}_T).$$

Using Bayes' theorem, the joint posterior density is

$$\begin{aligned} \pi\Big(\underset{\sim}{m}_0, \underset{\sim}{\beta}_i, \sigma^2, \tau_\mu, \tau_b, \rho, y_{0i} \mid \underset{\sim}{y}_i\Big) \propto & (\sigma^2)^{-(\ell T/2+1)} \exp\Bigg(-\frac{1}{2\sigma^2} \sum_{i=1}^{\ell} \|\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i\|^2 \Bigg) \\ & \times |\hat{\Sigma}_{(\tau_\mu, \tau_b, \rho)}|^{-\ell/2} \exp\Bigg(-\frac{1}{2} \sum_{i=1}^{\ell} (\underset{\sim}{\beta}_i - \underset{\sim}{m}_0)' \hat{\Sigma}^{-1}_{(\tau_\mu, \tau_b, \rho)} (\underset{\sim}{\beta}_i - \underset{\sim}{m}_0) \Bigg) \\ & \times \mathbf{1}\{0 < \tau_\mu < 1\} \, \mathbf{1}\{0 < \tau_b < 1\} \, \mathbf{1}\{-1 < \rho < 1\}. \end{aligned} \tag{11}$$

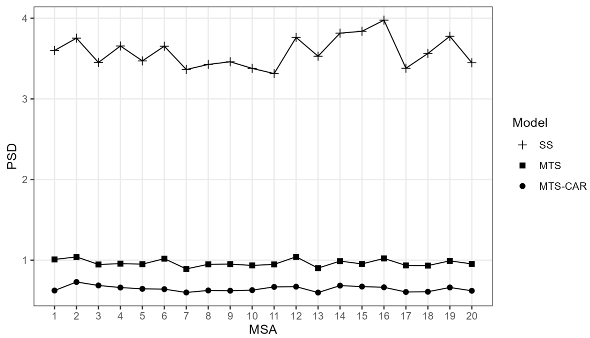
We describe the estimation of (11) via a Gibbs sampler in Appendix I (supplement)..

Although the model includes unit-specific parameters μ_i and b_i, it uses exchangeable hierarchical priors to share information across series through common hyperparameters. We regard this MTS specification as an intermediate benchmark between the single-series AR(1) baseline and the MTS-CAR model, which retains the same per-series likelihood but adds a CAR prior to encode cross-area dependence given the observed adjacency graph. This MTS model is the closest to the one described inNandram and Petruccelli (1997).

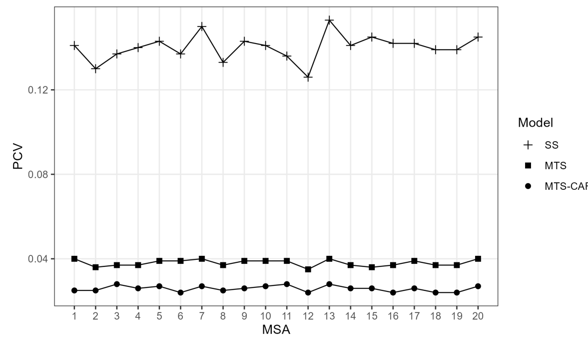
3.2.3. Empirical results

Here, we report empirical results for the baseline SS and MTS models and the proposed MTS-CAR model. We summarize posterior diagnostics, then compare one-step-ahead predictive accuracy and calibration across models using RMSE, PCV, PRMSE, bias, HPDI coverage, and relative error summaries. We analyze the data for 20 MSAs, but because of space, results for only 10 MSAs are shown in the tables for readability. The results are similar across all MSAs.

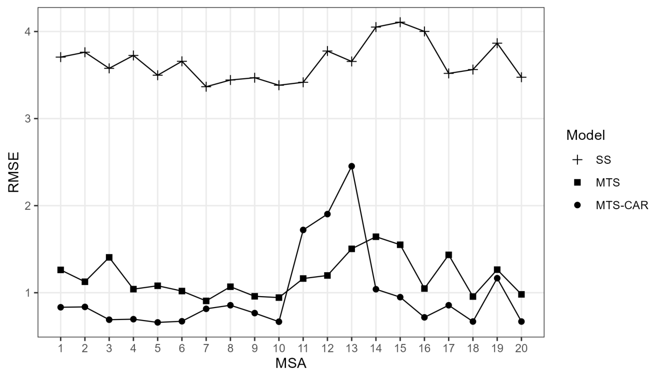
In Table 2, we report the posterior predictive summaries for the 10 MSAs under the three models: SS, MTS, and MTS-CAR. For comparison, the table also lists the actual value alongside the posterior predictive metrics. Across MSAs, the MTS-CAR and MTS models generally produce PM values close to the true value, with narrower HPDIs and smaller PSD and PCV, reflecting lower posterior variability and tighter uncertainty bands. Notably, MTS-CAR consistently attains the lowest PSD, BSTD, and PCV, indicating effective use of spatial structure, cross-series pooling, and limited between-area heterogeneity.



(a) PSD across MSAs



(b) PCV across MSAs



(c) RMSE across MSAs

Figure 1: Dispersion, variability, and error across MSAs.

Figure 1(a) and 1(b) illustrate the predictive dispersion and variability (PSD and PCV) across the three models for the 20 MSAs. Both MTS-CAR and MTS show low, stable dispersion, with MTS-CAR model achieving the lowest PSD and PCV across all the MSAs. The MTS model performs well, consistently better than SS, but typically trails the MTS-CAR model. These plots visually corroborate Table 2.

We present the RMSE, PB, and PARB as posterior diagnostic measures in Table 3. The MTS-CAR once again demonstrates strong predictive performance, outperforming the other models in a majority of the MSAs. Specifically, MTS-CAR achieves the lowest RMSE in 10 out of 10 MSAs reported, indicating that our proposed model consistently yields the most accurate forecasts. Additionally, MTS-CAR attains the lowest ARB in 5 out of 10 MSAs reported, with particularly low values like 0.002 in MSA 2, and 0.006 in MSA 4, 0.009 in MSA 3, and 0.027 in MSAs 7 and 8, demonstrating its ability to minimize relative prediction errors. The MTS model occasionally shows competitive ARB (*e.g.*, ARB of 0.003 in MSA 4, 0.004 in MSA 2, and 0.007 in MSA 3), attaining the lowest ARB in 4 out of the 10 MSAs reported. These results reaffirm the robustness of the MTS-CAR model in achieving both low prediction error and bias across a broad range of spatial units.

Figure 1(c) shows that our MTS-CAR model yields lower RMSE than MTS model and is far below SS for most MSAs, with especially large gains where SS errors are highest. The MTS-CAR model exhibits higher RMSE in a few units (*e.g.*, around MSAs 1113) than the MTS model, indicating heterogeneity in benefits. In MSAs where MTS-CAR does not uniformly outperform the baseline MTS model, the gains from spatial borrowing appear more

Table 2: Posterior predictive summary statistics

MSA	Model	True value	PM	PSD	BSTD	PCV	95% HPDI
1	SS	24.720	25.600	3.600	0.114	0.141	(18.177, 32.294)
	MTS		25.310	0.864	0.027	0.034	(23.631, 26.988)
	MTS-CAR		25.272	**0.624**	**0.020**	**0.025**	(24.104, 26.510)
2	SS	24.340	25.280	3.451	0.109	0.137	(19.062, 32.338)
	MTS		24.430	0.886	0.031	0.036	(22.749, 26.188)
	MTS-CAR		24.399	**0.687**	**0.022**	**0.028**	(23.074, 25.620)
3	SS	25.400	26.110	3.656	0.116	0.140	(19.066, 33.122)
	MTS		25.220	0.866	0.028	0.034	(23.725, 27.093)
	MTS-CAR		25.184	**0.661**	**0.021**	**0.026**	(23.901, 26.477)
4	SS	23.790	24.190	3.471	0.110	0.144	(17.692, 31.006)
	MTS		23.690	0.829	0.027	0.035	(22.018, 25.272)
	MTS-CAR		23.625	**0.645**	**0.020**	**0.027**	(22.208, 24.780)
5	SS	26.440	26.620	3.653	0.116	0.137	(20.224, 34.454)
	MTS		26.170	0.892	0.028	0.034	(24.121, 27.679)
	MTS-CAR		26.239	**0.641**	**0.020**	**0.024**	(25.014, 27.564)
6	SS	25.770	27.130	3.815	0.121	0.141	(19.898, 35.141)
	MTS		26.430	0.901	0.028	0.034	(24.638, 28.068)
	MTS-CAR		26.553	**0.685**	**0.022**	**0.026**	(25.329, 27.993)
7	SS	24.930	26.390	3.838	0.121	0.145	(19.167, 33.748)
	MTS		25.580	0.867	0.027	0.034	(24.024, 27.494)
	MTS-CAR		25.598	**0.673**	**0.021**	**0.026**	(24.355, 26.959)
8	SS	22.760	23.740	3.380	0.107	0.142	(17.586, 31.356)
	MTS		23.420	0.807	0.026	0.034	(21.787, 24.898)
	MTS-CAR		23.365	**0.606**	**0.019**	**0.026**	(22.015, 24.420)
9	SS	26.380	27.210	3.776	0.119	0.139	(20.417, 35.046)
	MTS		27.320	0.902	0.029	0.033	(25.719, 29.224)
	MTS-CAR		27.341	**0.662**	**0.021**	**0.024**	(26.019, 28.599)
10	SS	23.410	23.840	3.448	0.109	0.145	(17.336, 30.793)
	MTS		23.230	0.810	0.026	0.035	(21.634, 24.805)
	MTS-CAR		23.162	**0.621**	**0.020**	**0.027**	(21.964, 24.358)

Note: SS = Single-series, MTS = Multiple time series, MTS-CAR = Multiple time series with CAR prior, PM = Posterior Mean, PSD = Posterior Standard Deviation, BSTD = Bayesian Standard Deviation, PCV = Posterior Coefficient of Variation, HPDI = Highest Posterior Density Interval; Bolded figures denote the smallest values.

heterogeneous, likely reflecting weaker spatial coherence or stronger idiosyncratic dynamics in those regions. This behavior is expected and underscores that the benefits of adjacency-based pooling depend on the strength of underlying spatial similarity. Overall, the pattern corroborates the results in Table 3. Our MTS-CAR model generally improves accuracy relative to MTS model and substantially outperforms the SS model.

In Table 4, the MTS-CAR model frequently exhibited the lowest relative posterior error mean (PEM) and relative posterior error standard deviation (PESD) across almost all the MSAs, highlighting its capacity for precise and stable predictions. For instance, MTS-CAR achieved the lowest PEM in 7 out of the 10 MSAs reported, while MTS had the

Table 3: Prediction diagnostic

MSA	Model	RMSE	|PB|	ARB
1	SS	3.706	0.882	0.036
	MTS	1.262	0.757	0.024
	MTS-CAR	**0.833**	0.552	**0.022**
2	SS	3.577	0.941	0.039
	MTS	1.406	1.039	0.004
	MTS-CAR	**0.690**	0.059	**0.002**
3	SS	3.725	0.715	0.028
	MTS	1.041	0.409	**0.007**
	MTS-CAR	**0.696**	0.216	0.009
4	SS	3.498	0.430	0.018
	MTS	1.080	0.511	**0.003**
	MTS-CAR	**0.659**	0.135	0.006
5	SS	3.658	0.185	**0.007**
	MTS	0.931	0.267	0.010
	MTS-CAR	**0.672**	0.201	0.008
6	SS	4.051	1.360	0.053
	MTS	1.118	1.312	**0.026**
	MTS-CAR	**1.040**	0.783	0.030
7	SS	4.107	1.462	0.059
	MTS	1.086	0.653	**0.026**
	MTS-CAR	**0.949**	0.668	0.027
8	SS	3.519	0.980	0.043
	MTS	1.039	0.656	0.029
	MTS-CAR	**0.856**	0.605	**0.027**
9	SS	3.866	0.829	0.031
	MTS	1.301	0.938	0.036
	MTS-CAR	**1.167**	0.921	**0.029**
10	SS	3.474	0.429	0.018
	MTS	0.981	0.279	0.012
	MTS-CAR	**0.669**	0.248	**0.010**

Note: SS = Single series, MTS = Multiple time series, MTS-CAR = Multiple time series with CAR prior, RMSE = Posterior Root Mean Square Error; PB = Posterior Bias; ARB = Posterior Absolute Relative Bias; Bolded figures denote the smallest value.

lowest PEM in 3. MTS-CAR achieved the lowest PESD in all the 10 MSAs reported. The domination of our MTS-CAR model

suggest robust uncertainty quantification. Across the 20 MSAs, the MTS model offered moderate performance but was consistently outperformed by the MTS-CAR model. Overall, our proposed model provided stable and least dispersed prediction errors, and reliable interval estimates with relatively low error spread.

From a decision-theoretic perspective, tighter highest posterior density intervals and

Table 4: Posterior error summary statistics

MSA	Model	|PEM|	PESD	BSTD	95% HPDI
1	SS	0.031	0.152	0.005	(-0.242, 0.345)
	MTS	0.023	0.034	0.001	(-0.044, 0.088)
	MTS-CAR	**0.022**	**0.025**	0.001	(-0.024, 0.072)
2	SS	0.043	0.146	0.005	(-0.240, 0.330)
	MTS	0.004	0.035	0.001	(-0.063, 0.075)
	MTS-CAR	**0.002**	**0.027**	0.001	(-0.050, 0.054)
3	SS	0.019	0.145	0.005	(-0.280, 0.293)
	MTS	0.010	0.034	0.001	(-0.069, 0.066)
	MTS-CAR	**0.009**	**0.027**	0.001	(-0.054, 0.053)
4	SS	0.023	0.147	0.005	(-0.280, 0.293)
	MTS	**0.003**	0.033	0.001	(-0.074, 0.058)
	MTS-CAR	0.005	**0.026**	0.001	(-0.054, 0.048)
5	SS	0.005	0.139	0.004	(-0.236, 0.323)
	MTS	0.011	0.036	0.001	(-0.081, 0.058)
	MTS-CAR	**0.008**	**0.025**	0.001	(-0.057, 0.042)
6	SS	0.054	0.147	0.005	(-0.231, 0.333)
	MTS	0.036	0.036	0.001	(-0.051, 0.090)
	MTS-CAR	**0.031**	**0.027**	0.001	(-0.021, 0.085)
7	SS	0.057	0.146	0.005	(-0.234, 0.339)
	MTS	0.028	0.035	0.001	(-0.050, 0.089)
	MTS-CAR	**0.027**	**0.027**	0.001	(-0.021, 0.081)
8	SS	0.053	0.148	0.005	(-0.257, 0.317)
	MTS	0.026	0.032	0.001	(-0.038, 0.087)
	MTS-CAR	**0.024**	**0.024**	0.001	(-0.025, 0.070)
9	SS	0.041	0.136	0.004	(-0.212, 0.318)
	MTS	**0.037**	0.036	0.001	(-0.025, 0.114)
	MTS-CAR	0.038	**0.027**	0.001	(-0.015, 0.090)
10	SS	0.018	0.150	0.005	(-0.312, 0.296)
	MTS	**0.007**	0.032	0.001	(-0.073, 0.054)
	MTS-CAR	0.010	**0.025**	0.001	(-0.059, 0.037)

Note: SS = Single-series, MTS = Multiple time series; MTS-CAR = Multiple time series with CAR prior; PEM = relative Posterior error mean; PESD = relative Posterior error standard deviation; BSTD = Bayesian standard deviation; CV = Coefficient of variation; HPDI = Highest Posterior Density Interval; Bolded figures denote the smallest value.

reduced posterior dispersion are especially important in small-area forecasting, where decisions are often made under limited data and uncertainty directly affects resource allocation and planning. By borrowing strength across neighboring units through the prior, the MTS-CAR framework yields more concentrated posterior distributions without sacrificing interpretability. In addition, the collapsed Gibbs sampling strategy reduces posterior dependence among parameters, leading to improved mixing and effective sample sizes, consistent with modern Bayesian workflow principles that emphasize inferential efficiency alongside predictive performance.

Across multiple metrics, the results consistently demonstrate the advantages of our MTS-CAR model. The model repeatedly reduces prediction bias and relative error, reflecting the gains from pooling information across spatial units. The approach balances flexibility with efficiency and, relative to the SS and the MTS baselines, delivers more stable and accurate forecasts. Overall, the findings reinforce the value of hierarchical Bayesian modeling for small-area estimation, highlighting the complementary benefits of spatially adaptive pooled coefficient structures.

4. Concluding remark

This paper tackles forecasting in spatiotemporal settings - data evolving over time and across locations - where each location often has limited observations, as is common in local economic analyses. We propose a new multiple series model (MTS-CAR) that embeds spatially informed priors, enabling it to capture local earnings dynamics while accounting for spillover effects from neighboring regions.

Our MTS-CAR specification places a conditional autoregressive (CAR) prior on the unit-specific intercepts and extends this structure to the coefficient blocks via a shared spatial precision $(D - \rho W)$, while the autoregressive coefficients are hierarchically pooled around global means. This setup yields flexible, region-specific dynamics that borrow strength across neighboring areas without over-parameterization. We build on the hierarchical panel AR framework of Nandram and Petruccelli (1997) by retaining per-unit AR likelihoods and hierarchical pooling, but replacing exchangeable (global) pooling with adjacency-aware CAR priors on the unit-specific intercept and AR-slope deviations, thereby introducing explicit spatial dependence and coefficient-level regularization (spatially adaptive shrinkage). In contrast to spatial-econometric specifications with a lagged dependent variable in the likelihood (*e.g.*, SAR/Spatial autoregressive with autoregressive disturbances (SARAR); see LeSage and Pace (2009), building on Cliff and Ord (1973), and Anselin (1988)), our model introduces spatial heterogeneity through the prior, achieving adjacency-aware pooling while retaining per-series univariate likelihoods.

To test the effectiveness of our model, we applied it to panel data on earnings across multiple Metropolitan Statistical Areas (MSAs), and compared the results to the baseline models. The empirical results demonstrate that the MTS-CAR framework reliably improves predictive performance in small-area, multi-series settings by borrowing strength across spatially related units. Relative to the autoregressive single series (SS) and multiple time series (MTS) baselines, MTS-CAR delivers tighter uncertainty quantification, lower dispersion, and reduced error, indicating both greater stability and accuracy. The model's out-of-sample performance was impressive. We saw a median reduction of 30.8% (maximum 50.9%) in the root mean square forecast errors relative to the MTS baseline across the 20 areas. These gains arise from explicit spatial pooling that preserves unit-level heterogeneity while exploiting shared structure. Beyond point accuracy, the model concentrates posterior mass more effectively, yielding sharper, decision-relevant intervals. Robustness checks suggest conclusions are not unduly sensitive to reasonable choices of the spatial weights matrix W and prior hyperparameters. Taken together, the evidence supports hierarchical Bayesian modeling with spatial priors as a practical and principled approach for small-area estimation and forecasting.

Conceptually, the proposed MTS-CAR framework differs from likelihood-based spatial autoregressive models by introducing dependence entirely through the prior, rather than through the observation equation. By operating at the coefficient level and leveraging collapsed posterior updates, the model delivers interpretable dynamics, efficient inference, and decision-relevant uncertainty quantification for small-area forecasting.

While our sensitivity checks show robustness to reasonable choices of the spatial weights W and prior hyperparameters, the current specification relies on a fixed W and a time-invariant ρ; fully dynamic spatial dependence and alternative proximity measures are left for future work. Crucially, the tighter intervals and lower dispersion we obtained are decision-useful. They support clearer budgeting ranges, more credible risk bounds, and sharper targeting for regional labor interventions. The framework is readily transferable to other small-area panels, such as health, education, or local price dynamics, where spillovers and sparse unit-level samples coexist without sacrificing interpretability. Natural extensions include time-varying spatial strength ρ_t, domain-specific W (*e.g.*, commuting, trade, or contiguity hybrids), inclusion of exogenous covariates, principled missing-data handling, and scalable inference for larger ℓ and longer horizons.

Moreover, while we focus on multiple univariate series with adjacency-aware pooling, one avenue for future work is to allow higher-order temporal dynamics while retaining adjacency-aware pooling. Any univariate $AR(p)$ (and, more generally, VAR(p)) can be written in companion form as an $AR(1)$ in an expanded state, so stationarity reduces to an eigenvalue check: the spectral radius of the companion (first-order) coefficient matrix must be strictly less than one. This makes it straightforward to monitor or enforce stationarity during posterior simulation (*e.g.*, reject/transform draws that violate the constraint) without altering the CAR prior structure. See Lütkepohl (2005) for the standard treatment of companion representations and stationarity conditions. We leave this multivariate extension to future research.

More broadly, our evidence supports hierarchical Bayesian small-area methods Rao and Molina (2015), Bayesian shrinkage for univariate time-varying parameter models (Belmonte *et al.*, 2014), and Bayesian pooling across many univariate AR processes (Rao and Yu, 1994; Nandram and Petruccelli, 1997) as practical tools for forecasting in spatially interconnected, data-sparse settings.

Acknowledgments

The authors are grateful for the invitation to present a paper in the special issue, Recent Advances in Bayesian Statistics and Machine Learning, under the topic, Bayesian methods for time series, spatial statistics, and complex networks.

Conflict of interest

The authors declare no financial or non-financial conflicts of interest related to this work.

Supplementary materials

All R codes and notes for reproducing the analyses of the MTS and MTS-CAR models in this paper and the supplementary material are available at: `https://github.com/haolayinka/MTS-CAR-Model`

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.

Belmonte, M. A., Koop, G., and Korobilis, D. (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, **33**, 80–94.

Binder, D. A. and Dick, J. P. (1989). Modelling and estimation for repeated surveys. *Survey Methodology*, **15**, 29–45.

Cadonna, A., Kottas, A., and Prado, R. (2019). Bayesian spectral modeling for multiple time series. *Journal of the American Statistical Association*, **114**, 1838–1853.

Cliff, A. D. and Ord, J. K. (1973). *Spatial Autocorrelation*. Pion Limited, London.

Desjardins, M. R., Hohl, A., and Delmelle, E. M. (2020). Space-time conditional autoregressive modeling to estimate COVID-19 risk. *International Journal of Environmental Research and Public Health*, **17**, 545.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269–277.

Gefang, D., Hall, S. G., and Tavlas, G. S. (2025). Fast two-stage variational Bayesian approach to estimating panel spatial autoregressive models with unrestricted spatial weights matrices. *Spatial Economic Analysis*, 1–26.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, **4**, 11–15.

Girolimetto, D. and Di Fonzo, T. (2024). Insights into regression-based cross-temporal forecast reconciliation. In *Scientific Meeting of the Italian Statistical Society*, pages 119–125, Cham. Springer Nature Switzerland.

Golub, G. H. and Van Loan, C. F. (2013). *Matrix Computations*. Johns Hopkins University Press, 4 edition.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Review*, **23**, 53–60.

Jin, X., Carlin, B. P., and Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics*, **61**, 950–961.

Keefe, M. J., Ferreira, M. A. R., and Frank, C. T. (2019). Objective Bayesian analysis for Gaussian hierarchical models with intrinsic conditional autoregressive priors. *Bayesian Analysis*, **14**, 181–209.

Lawson, A. B. (2018). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Chapman and Hall/CRC, 3 edition.

Lee, D., Mitchell, R., and Sahu, S. K. (2014). A Bayesian localized conditional autoregressive model for estimating the health effects of air pollution. *Biometrics*, **70**, 419–429.

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

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

Manandhar, B. and Nandram, B. (2021). Hierarchical Bayesian models for continuous and positively skewed data from small areas. *Communications in Statistics-Theory and Methods*, **50**, 944–962.

Manandhar, B. and Nandram, B. (2025). Hierarchical Bayesian models for small area estimation with GB2 distribution. *Journal of Applied Statistics*, **52**, 1–30.

Mozdzen, A., Cremaschi, A., Cadonna, A., Guglielmi, A., and Kastner, G. (2022). Bayesian modeling and clustering for spatio-temporal areal data: An application to Italian unemployment. *Spatial Statistics*, **52**, 100715.

Müller, U. K. and Watson, M. W. (2025). Forecasting related time series. Technical report, Department of Economics, Princeton University. Working paper, draft dated October 22, 2025.

Nandram, B. and Petruccelli, J. D. (1997). A Bayesian analysis of autoregressive time series panel data. *Journal of Business & Economic Statistics*, **15**, 328–334.

Nie, Y., Nguyen, N. H., Sinthong, P., and Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*.

Nieto-Barajas, L. E. and Quintana, F. A. (2016). A bayesian non-parametric dynamic AR model for multiple time series analysis. *Journal of Time Series Analysis*, **37**, 675–689.

Orozco-Acosta, E., Arab, A., Cruz, M., and Finley, A. (2023). A scalable approach for short-term disease forecasting in space and time. *Biometrical Journal*, **65**, 2200096.

Prado, R., Ferreira, M. A. R., and West, M. (2021). *Time Series: Modeling, Computation, and Inference*. Chapman and Hall/CRC, 2 edition.

Rao, J. N. K. and Molina, I. (2015). *Small Area Estimation*. John Wiley & Sons, 2nd edition.

Rao, J. N. K. and Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *Canadian Journal of Statistics*, **22**, 511–528.

Rushworth, A., Lee, D., and Mitchell, R. (2014). A spatio-temporal model for estimating the long-term effects of air pollution on health. *Spatial and Spatio-temporal Epidemiology*, **10**, 29–38.

Sahu, S. K. (2021). *Bayesian Modeling of Spatio-Temporal Data with R*. Chapman and Hall/CRC.

Sherman, J. and Morrison, W. J. (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, **21**, 124–127.

Shoesmith, G. L. (2013). Space–time autoregressive models and forecasting national, regional and state crime rates. *International Journal of Forecasting*, **29**, 191–201.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, **48**, 1–48.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer, New York, 2nd edition.

Woodbury, M. A. (1950). Inverting modified matrices. Technical report, Department of Statistics, Princeton University.

Yu, B., Yin, H., and Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*.