

Bayesian inference for dynamic spatial quantile models with interactive effects *

Tomohiro Ando ^{†1}, Jushan Bai², Kunpeng Li³, and Yong Song⁴

¹The University of Melbourne Melbourne Business School

²Columbia University

³Capital University of Economics and Business

⁴University of Melbourne

June 9, 2025

Abstract

This paper proposes a dynamic spatial panel quantile model that accounts for unobserved heterogeneity. The Bayesian MCMC algorithm developed here introduces several novel features tailored to the challenges of high-dimensional dynamic spatial quantile models. Key innovations include quantile randomisation, which enables conditional conjugacy and facilitates efficient sampling, and a new Gibbs sampler specifically designed to handle structural spatial parameters. We also establish the Bayesian consistency of the proposed estimation method. Finally, we demonstrate the practical utility of the approach through an application to the quantile comovement structure of the Australian gasoline market.

Keywords: Dynamic panel, endogeneity, factor models, heterogenous spatial effects, high dimensional data.

JEL classification: C31, C33, E44

*The paper has benefited from constructive comments from participants at Australian spatial econometrics and statistics workshop 2023, held at Monash university. This research was supported by University of Melbourne Research Computing Services and the Petascale Campus Initiative. It is also supported by the Australian Research Council Discovery Grants DP230100959 and DP240101009.

[†]A part of this work was completed when TA was visiting Monash University. TA would like to thank their generous supports during the visit.

1 Introduction

Spatial panel data models explore spatial interactions across individuals or regions over time. These models have found wide-ranging applications in fields such as housing economics (e.g., [Beenstock and Felsenstein \(2015\)](#)), marketing (e.g., [Hunneman et al. \(2022\)](#)), and urban economics (e.g., [Glaser et al. \(2022\)](#)). In this paper, we introduce a novel dynamic spatial panel quantile model with interactive effects. This model simultaneously accounts for spatial dependence, temporal dynamics, unobserved heterogeneity, and heterogeneous effects across quantiles—features that are often observed in economic and social data but rarely modeled together.

There is a rich literature on linear panel models with interactive effects (e.g., [Bai \(2009\)](#); [Bai and Li \(2014\)](#); [Ando and Bai \(2017\)](#); [Hallin and Liška \(2007\)](#); [Moon and Weidner \(2015\)](#); [Pesaran \(2006\)](#); [Stock and Watson \(2002\)](#); [Lu and Su \(2016\)](#)) and linear spatial panel models (e.g., [Aquaro et al. \(2021\)](#); [Baltagi \(2011\)](#); [Bai and Li \(2021\)](#); [Kelejian and Prucha \(2004\)](#); [Lee \(2004\)](#); [Lin and Lee \(2010\)](#); [Li \(2017\)](#); [Lu \(2017\)](#); [Qu and Lee \(2015\)](#); [Reich et al. \(2011\)](#); [Shi and Lee \(2017\)](#); [Yu et al. \(2008\)](#)). However, research on quantile regression (see [Koenker and Bassett \(1978\)](#)) in spatial and high-dimensional panel settings remains limited, especially for models that incorporate both interactive effects and dynamic structures. Recent work by [Ando and Bai \(2020\)](#) extended quantile methods to panel data with interactive effects, capturing quantile co-movements while allowing for heterogeneous slope coefficients. However, their model does not accommodate dynamic and spatial dependencies in quantiles.

Spatial interactions and temporal dynamics are fundamental features of economics and the social sciences. Spatial dependence may arise from peer effects, spillover effects, or imitative behavior (see [Manski \(1993\)](#); [Anselin \(1988\)](#), among others), while temporal dependence often reflects persistence in individual or group behavior over time. To capture these complex features, we propose a dynamic spatial quantile panel model with interactive effects. This framework facilitates quantile-specific analysis of both spatial and temporal dependence, accommodates heterogeneous slope coefficients, and incorpo-

rates latent factor structures. Due to the model’s high-dimensional parameter space—driven by the dynamic structure and large number of cross-sectional units—we develop a new Bayesian Markov Chain Monte Carlo (MCMC) estimation method tailored to this setting.

Methodological challenges: Several methodological challenges arise in the estimation process. First, because the quantile function is defined implicitly in terms of its coefficients, a conventional Gibbs sampler is not directly applicable. To overcome this, we introduce randomised quantile dynamics that create conditional conjugacy, allowing for efficient sampling. Second, the spatial autoregressive parameter is structural in nature and does not permit a standard Gibbs step. Inspired by methods in structural vector autoregression, we construct a bimodal mixture distribution to incorporate this parameter within our MCMC framework. Finally, to alleviate the computational burden from high-dimensional matrix operations, we apply a breadth-first search algorithm from graph theory to block-diagonalise the spatial matrix, thereby significantly reducing computational complexity. These methodological innovations enable our MCMC algorithm to efficiently estimate large-scale dynamic spatial quantile panel models with latent factor structures.

Theoretical Challenges: To establish the validity of our Bayesian approach, we prove Bayesian consistency—a theoretical result not previously achieved for panel data models with interactive effects. This requires overcoming several substantial hurdles: the dynamic structure of the model, high-dimensional incidental parameters from latent factors and their loadings, the nonsmooth nature of the quantile objective function, the nonlinearity introduced by the spatial term, and the rotational indeterminacy of the factor structure. Although recent work (e.g., [Ando and Bai \(2020\)](#); [Chen et al. \(2021\)](#)) addresses some related issues, their results are not applicable to our setting. Consequently, we first derive new asymptotic properties within a frequentist framework (see Remark 7), which we then build upon to develop our Bayesian theory. Even with these foundations, our model demands additional theoretical innovation to rigorously establish Bayesian consistency.

Our contributions are summarised as follows. First, we introduce a dynamic spa-

tial panel quantile model with interactive effects, designed to accommodate key features such as spatial dependence and temporal dynamics. This specification enhances the model’s flexibility in capturing complex data patterns. Second, we develop a novel Bayesian estimation procedure tailored to high-dimensional parameter spaces. Third, we establish several asymptotic properties, including Bayesian consistency. Finally, we demonstrate the utility of our framework through an empirical analysis of the Australian gasoline market.

The paper is organised as follows. Section 2 introduces a new dynamic spatial panel quantile model with interactive fixed effects. Section 3 outlines the proposed Bayesian MCMC estimation procedure. Section 4 establishes the Bayesian consistency of the method. In Section 5, we apply the proposed approach to the Australian gasoline market. Section 6 concludes the paper. To conserve space, all technical proofs are provided in the online supplementary material. The supplementary document also includes Monte Carlo simulation results, which demonstrate the effectiveness of the proposed estimation procedure.

Notations: Let $\|A\| = [\text{tr}(A'A)]^{1/2}$ be the Frobenius norm of matrix A , where “tr” denotes the trace of a square matrix, and let $\|A\|_2$ be its spectrum norm (the largest singular value of A). In addition, for any $N \times N$ matrix $\|A\|_1$ is defined as $\|A\|_1 = \max_{1 \leq j \leq N} \sum_{i=1}^N |a_{ij}|$ where a_{ij} is the (i, j) -th element of A . Similarly, $\|A\|_\infty = \max_{1 \leq i \leq N} \sum_{j=1}^N |a_{ij}|$. For sequences a_n and b_n , the notation $a_n \lesssim b_n$ means $a_n = O(b_n)$, that is, there exists $C > 0$ and for all n large enough, $a_n \leq Cb_n$. We write $c_n = O_p(d_n)$ if c_n/d_n is stochastically bounded, and $c_n = o_p(d_n)$ if c_n/d_n converges to zero in probability.

2 Dynamic spatial quantile models with interactive effects

Suppose that, for the i -th unit ($i = 1, \dots, N$) at time t ($t = 1, \dots, T$), its response y_{it} is observed together with a set of p explanatory variables $\{x_{it,1}, \dots, x_{it,p}\}$. We consider the τ -th quantile function of y_{it} by jointly modeling spatial effects, time effects and common shocks. To capture these effects simultaneously, we define the τ -th quantile function of

y_{it} as

$$\begin{aligned}
& Q_{y_{it}} \left(\tau | X_t, F_{t,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau \right) \\
& \equiv \rho_{i,\tau} \sum_{i \neq j, j=1}^N w_{ij} Q_{y_{jt}} \left(\tau | X_t, F_{t,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau \right) \\
& \quad + \delta_{i,\tau} \sum_{i \neq j, j=1}^N w_{ij} Q_{y_{j,t-1}} \left(\tau | X_{t-1}, F_{t-1,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau \right) \\
& \quad + \gamma_{i,\tau} Q_{y_{i,t-1}} \left(\tau | X_{t-1}, F_{t-1,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau \right) + \mathbf{x}'_{it} \mathbf{b}_{i,\tau} + \mathbf{f}'_{t,\tau} \boldsymbol{\lambda}_{i,\tau} \tag{1}
\end{aligned}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. Here w_{ij} ($i = 1, 2, \dots, N; j = 1, 2, \dots, N$) are pre-specified spatial weights with $w_{ii} = 0$, $\rho_{i,\tau}$ and $\delta_{i,\tau}$ are the heterogeneous spatial parameters capturing the strength of the spillover effects, the coefficients $\gamma_{i,\tau}$ are the heterogeneous temporal parameters, $\mathbf{x}_{it} = (1, x_{it,1}, \dots, x_{it,p})'$ is $(p + 1)$ -dimensional vector of explanatory variables; $B_\tau = (\mathbf{b}_{1,\tau}, \mathbf{b}_{2,\tau}, \dots, \mathbf{b}_{N,\tau})'$, $\mathbf{b}_{i,\tau} = (b_{i,0,\tau}, b_{i,1,\tau}, \dots, b_{i,p,\tau})'$ is a $(p + 1)$ -dimensional vector of regression coefficients; $\mathbf{f}_{t,\tau} = (f_{t1,\tau}, \dots, f_{tr_\tau,\tau})'$ is r_τ -dimensional unobservable common factors; $\boldsymbol{\lambda}_{i,\tau} = (\lambda_{i1,\tau}, \dots, \lambda_{ir_\tau,\tau})'$ is r_τ -dimensional vector of factor loadings; X_t and $F_{t,\tau}$ are information on the explanatory variables and the common factors up to time t . The exact expression of $\mathbf{x}'_{it} \mathbf{b}_{i,\tau}$ is

$$\mathbf{x}'_{it} \mathbf{b}_{i,\tau} = \sum_{k=1}^p x_{it,k} b_{ik,\tau} + G_{i,e_{it}}^{-1}(\tau), \tag{2}$$

where e_{it} is the idiosyncratic error term and $G_{i,e_{it}}^{-1}(\tau)$ is the τ -th quantile point of e_{it} with $G_{i,e_{it}}(\cdot)$ being the cumulative distribution function of e_{it} . Thus, the τ -th quantile of the idiosyncratic error $G_{i,e_{it}}^{-1}(\tau)$, which depends only on i and τ , is absorbed by the term $\mathbf{x}'_{it} \mathbf{b}_{i,\tau}$ since the first element of \mathbf{x}_{it} is 1. We assume that e_{it} is identically distributed over t while its distribution may vary over i .

The preceding quantile function (1) is associated with the following data-generating process, provided that the right-hand side of the equation is an increasing function of u_{it} ,

$$y_{it,u_{it}} = \rho_{i,u_{it}} \sum_{j=1}^N w_{ij} y_{jt,u_{it}} + \gamma_{i,u_{it}} y_{i,t-1,u_{it}} + \delta_{i,u_{it}} \sum_{j=1}^N w_{ij} y_{j,t-1,u_{it}} + \mathbf{x}'_{it} \mathbf{b}_{i,u_{it}} + \mathbf{f}'_{t,u_{it}} \boldsymbol{\lambda}_{i,u_{it}},$$

where $y_{it,u_{it}}$ is $\tau = u_{it}$ -th quantile and u_{it} are i.i.d. $U(0, 1)$. The coefficient of the constant regressor absorbs the error term. As seen, we integrate spatial interactions and temporal

dynamics into a model, which allows for both contemporaneous and dynamic spatial effects, thereby enabling the capture of temporal spillover effects and peer influences in the spatial domain.

Remark 1 *Koenker and Xiao (2006) considered autoregressive quantile model in the context of univariate time series. Their quantile function is expressed as a weighted sum of past observed values of the response variable. While their model is regarded as autoregressive in this sense, the quantile function itself is not autoregressive. In contrast, our quantile function in (1) includes a weighted sum of past quantile functions, making the quantile itself autoregressive, or dynamic.*

Define the $N \times N$ matrix $S(\boldsymbol{\rho}_\tau) \equiv (I - \boldsymbol{\rho}_\tau W)^{-1}$, where $\boldsymbol{\rho}_\tau = \text{diag}(\rho_{1,\tau}, \dots, \rho_{N,\tau})$, and $W = [w_{ij}]$ is the $N \times N$ spatial weights matrix. Also, we define the $N \times N$ matrix $A(\boldsymbol{\rho}_\tau, \boldsymbol{\delta}_\tau, \boldsymbol{\gamma}_\tau) \equiv (I - \boldsymbol{\rho}_\tau W)^{-1}(\boldsymbol{\gamma}_\tau + \boldsymbol{\delta}_\tau W)$ where $\boldsymbol{\delta}_\tau = \text{diag}(\delta_{1,\tau}, \dots, \delta_{N,\tau})$, $\boldsymbol{\gamma}_\tau = \text{diag}(\gamma_{1,\tau}, \dots, \gamma_{N,\tau})$. Stack the quantile functions over cross sections by defining

$$\mathbf{Q}_t \left(X_t, F_{t,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau \right) = \begin{bmatrix} Q_{y_{1t}}(\tau | X_t, F_{t,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau) \\ \vdots \\ Q_{y_{Nt}}(\tau | X_t, F_{t,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau) \end{bmatrix}$$

By recursive substitution, model (1) can be rewritten as

$$\mathbf{Q}_t \left(X_t, F_{t,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau \right) = \sum_{h=0}^{t-1} P_h(\boldsymbol{\rho}_\tau, \boldsymbol{\delta}_\tau, \boldsymbol{\gamma}_\tau) \begin{bmatrix} \mathbf{x}'_{1,t-h} \mathbf{b}_{1,\tau} + \mathbf{f}'_{t-h,\tau} \boldsymbol{\lambda}_{1,\tau} \\ \vdots \\ \mathbf{x}'_{N,t-h} \mathbf{b}_{N,\tau} + \mathbf{f}'_{t-h,\tau} \boldsymbol{\lambda}_{N,\tau} \end{bmatrix},$$

where the $N \times N$ matrix $P_h(\boldsymbol{\rho}_\tau, \boldsymbol{\delta}_\tau, \boldsymbol{\gamma}_\tau)$ is given as

$$P_h(\boldsymbol{\rho}_\tau, \boldsymbol{\delta}_\tau, \boldsymbol{\gamma}_\tau) \equiv A(\boldsymbol{\rho}_\tau, \boldsymbol{\delta}_\tau, \boldsymbol{\gamma}_\tau)^h S(\boldsymbol{\rho}_\tau). \quad (3)$$

We will impose restrictions on (3) to ensure the sum to be well defined (referred to as stationarity). In addition, we assume $\mathbf{Q}_t(\cdot) = 0$ when $t = 0$. The effect of the initial condition is generally negligible if T is large. Thus, an alternative expression of (1) is

$$Q_{y_{jt}} \left(\tau | X_t, B_\tau, F_{t,\tau}, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau \right) = \sum_{h=0}^{t-1} \sum_{k=1}^N p_{jk,h}(\boldsymbol{\rho}_\tau, \boldsymbol{\delta}_\tau, \boldsymbol{\gamma}_\tau) \left(\mathbf{x}'_{k,t-h} \mathbf{b}_{k,\tau} + \mathbf{f}'_{t-h,\tau} \boldsymbol{\lambda}_{k,\tau} \right), \quad (4)$$

where $p_{ij,h}(\boldsymbol{\rho}_\tau, \boldsymbol{\delta}_\tau, \boldsymbol{\gamma}_\tau)$ is the (i, j) th element of $P_h(\boldsymbol{\rho}_\tau, \boldsymbol{\delta}_\tau, \boldsymbol{\gamma}_\tau)$ in (3).

To eliminate the rotational indeterminacy of the common factor structure, we need to impose a restriction on $F_\tau = (\mathbf{f}_{1,\tau}, \dots, \mathbf{f}_{T,\tau})'$ and $\Lambda_\tau = (\boldsymbol{\lambda}_{1,\tau}, \dots, \boldsymbol{\lambda}_{N,\tau})'$. For example, [Bai and Li \(2013\)](#) imposed the followings

$$\frac{1}{T}F_\tau'F_\tau = I_{r_\tau} \quad \text{and} \quad \frac{1}{N}\Lambda_\tau'\Lambda_\tau = D_{r_\tau}, \quad (5)$$

where I_{r_τ} is an $r_\tau \times r_\tau$ identity matrix, and D_{r_τ} is a diagonal matrix whose diagonal elements are distinct and are arranged in a descending order. We refer to [Bai and Ng \(2013\)](#) for alternative restrictions on the common factor structure.

We have to estimate the unknown parameters $\boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau, B_\tau, \Lambda_\tau$, and F_τ simultaneously. Let $\vartheta_\tau = \{\boldsymbol{\rho}_\tau, \boldsymbol{\delta}_\tau, \boldsymbol{\gamma}_\tau, B_\tau, \Lambda_\tau, F_\tau\}$. The frequentist estimator can be obtained as the minimiser of the following objective function

$$\ell_\tau(Y|X, \vartheta_\tau) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T q_\tau(y_{it} - Q_{y_{it}}(\tau|X_t, \vartheta_\tau)) \quad (6)$$

where $Q_{y_{it}}(\tau|X_t, \vartheta_\tau) \equiv Q_{y_{it}}(\tau|X_t, F_{t,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau)$ is defined in (1), $q_\tau(u) = u(\tau - I(u \leq 0))$ is the quantile loss function, $Y \equiv \{y_{it}|i = 1, \dots, N, t = 1, \dots, T\}$ and $X \equiv \{\mathbf{x}_{it}|i = 1, \dots, N, t = 1, \dots, T\}$.

3 Bayesian estimation

In this paper, we employ Bayesian MCMC sampling procedure to produce posterior samples from the posterior distribution

$$\pi(\vartheta_\tau|X, Y) \propto \exp(-\ell_\tau(Y|X, \vartheta_\tau)) \pi(\vartheta_\tau), \quad (7)$$

where $\pi(\vartheta_\tau)$ is a prior density of the parameter ϑ_τ . Details of $\pi(\vartheta_\tau)$ are given below.

The quantile structure in (4) is a highly nonlinear function of parameters $\boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau$ and B_τ . Such nonlinearity costs the conditional conjugacy for inference, rendering the ideal Gibbs sampling infeasible. Simple methods such as the Metropolis-Hastings might be very inefficient due to the large number of parameters. For example, in our application,

the dimension of the parameter space is greater than 44,000. In this paper, we propose to randomise the quantile equation to adapt to the Bayesian inference framework. In particular, we set

$$\varepsilon_{it,\tau} = y_{it} - Q_{y_{it}}(\tau|X_t, \vartheta_\tau), \quad (8)$$

$$Q_{it,\tau} = \rho_{i,\tau} \sum_{j=1, j \neq i}^N w_{ij} Q_{jt,\tau} + \gamma_{i,\tau} Q_{i,t-1,\tau} + \delta_{i,\tau} \sum_{j=1, j \neq i}^N w_{ij} Q_{j,t-1,\tau} + \mathbf{x}'_{it} \mathbf{b}_{i,\tau} + \mathbf{f}'_{t,\tau} \boldsymbol{\lambda}_{i,\tau} + e^q_{it,\tau}. \quad (9)$$

In the Bayesian framework, it is commonly assumed that $\varepsilon_{it,\tau}$ in (8) has an Asymmetric Laplace distribution (ALD): $p(\varepsilon_{it,\tau} | \tau, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp(-\frac{q_\tau(\varepsilon_{it,\tau})}{\sigma})$. The Gibbs sampling method of [Kozumi and Kobayashi \(2011\)](#) uses a scaled mixture representation of the ALD. Under this representation, the conditional density of y_{it} , given the auxiliary variable V_{it} , is expressed as normal:

$$p(y_{it} | X_t, \vartheta_\tau, V_{it}, \sigma) \propto \sqrt{\frac{\tau(1-\tau)}{2\sigma^2 V_{it}}} \exp\left(-\frac{(y_{it} - Q_{y_{it}}(\tau|X_t, \vartheta_\tau) - \frac{1-2\tau}{\tau(1-\tau)}\sigma V_{it})^2}{\frac{4}{\tau(1-\tau)}\sigma^2 V_{it}}\right), \quad (10)$$

where $V_{it} \sim \text{Exp}(1)$ and σ is the temperature parameter to facilitate Bayesian inference.

We denote $Q_{y_{it}}(\tau|X_t, \vartheta_\tau)$ as $Q_{it,\tau}$ in (9), and $e^q_{it,\tau} \sim N(0, \sigma_{q,\tau}^2)$. Note that $e_{it,\tau}$ in (2) and $e^q_{it,\tau}$ in (9) are different. The introduction of $e^q_{it,\tau}$ is the randomisation that allows to treat the quantile values as a latent variable to augment the parameter space. Without $e^q_{it,\tau}$, the $Q_{it,\tau}$ is a deterministic function of ϑ_τ in (9) and hence cannot be treated as conditioning parameters. Consequently, the pseudo posterior kernel (7) will be a complicated function of ϑ_τ , making posterior computation burdensome, particularly in high-dimensional settings. By introducing the randomised component $e^q_{it,\tau}$, the process $Q_{it,\tau}$ gains variability and can be considered part of the augmented parameter space. This randomisation enables efficient Gibbs sampling by allowing the conditional distributions to take tractable forms. In this way, equation (9) plays a key role in making Gibbs sampling computationally feasible. See also our discussion in Remark 3.

3.1 Prior setting

We apply the same prior to the parameters for each quantile τ . So ignore τ for notational simplicity. The parameter space includes:

1. A normal prior is used for the lag parameter $\gamma_i \sim N(m_\gamma, h_\gamma^{-1})$, $i = 1, 2, \dots, N$.
2. A normal prior is used for the spatial lag parameter $\delta_i \sim N(m_\delta, h_\delta^{-1})$, $i = 1, 2, \dots, N$.
3. A multivariate normal prior is used for $\mathbf{b}_i \sim N(\mathbf{m}_b, H_b^{-1})$, $i = 1, 2, \dots, N$.
4. A normal prior is used for the spatial parameter $\rho_i \sim N(m_\rho, h_\rho^{-1})$, $i = 1, 2, \dots, N$.
5. For the common factor, we assume a stationary AR(1) process as $f_{j,t} = \phi_j f_{j,t-1} + e_{j,t}^f$, for $j = 1, \dots, r$, where $e_{j,t}^f \sim N(0, 1)$ with a unit variance for identification purpose. Each $e_{j,t}^f$ is independent over j and t . Also, we assume the initial condition $f_{j,1} \sim N(0, h_f^{-1})$, for $j = 1, \dots, r$.
6. For the autoregressive coefficient, we set $\phi_j \sim U(-1, 1)$, $j = 1, \dots, r$.
7. For the factor loading λ_{ij} , we use a normal prior: $\lambda_{ij} \sim N(0, h_\lambda^{-1})$. The first $r \times r$ block is a lower triangular matrix such that $\lambda_{ii} \sim N(0, h_\lambda^{-1})\mathbf{1}(\lambda_{ii} > 0)$ for $i = 1, \dots, r$, and $\lambda_{ij} = 0$ if $j > i$.
8. Gamma prior $\sigma \sim G(v_\sigma, s_\sigma)$ is used for σ in the asymmetric Laplace error term.

Remark 2 In addition to ϑ_τ in the model (1), the Bayesian parameter space also has auxiliary variables $\vartheta'_\tau = \{Q_\tau, V, \phi, \sigma\}$, where V and Q_τ are the collections of V_{it} in (10) and $Q_{it,\tau}$, respectively. Denote the corresponding posterior as

$$p(\vartheta_\tau, \vartheta'_\tau \mid X, Y) \propto p(Y \mid X, \vartheta_\tau, V, \sigma) \pi(\vartheta_\tau) \pi(\vartheta'_\tau) \quad (11)$$

where $p(Y \mid X, \vartheta_\tau, V, \sigma) = \prod_{i=1}^N \prod_{t=1}^T p(y_{it} \mid X_t, \vartheta_\tau, V_{it}, \sigma)$ with $p(y_{it} \mid X_t, \vartheta_\tau, V_{it}, \sigma)$ given in (10), and $\pi(\vartheta'_\tau)$ is a prior density of ϑ'_τ . The concentrated posterior for comparison with the frequentist parameter set is defined as $p(\vartheta \mid X, Y) = \int p(\vartheta, \vartheta' \mid X, Y) d\vartheta'$. Although $p(\vartheta \mid X, Y)$ here and $\pi(\vartheta \mid X, Y)$ in (7) are different, Proposition 1 in Section 3.2 below ensures that our MCMC posterior samples generated from $p(\vartheta \mid X, Y)$ can be regarded as those generated from $\pi(\vartheta \mid X, Y)$ through the importance sampling procedure.

3.2 MCMC sampling procedure

We briefly describe the procedure in this section, with detailed techniques available in the Appendix. The computational challenge of our model arises from three main factors. First, the high dimensionality of the quantile values and their associated dynamics require a significant number of large matrix inversions. Second, due to the substantial heterogeneity in the model, generic methods, such as Metropolis-Hastings or particle filters, face high computational cost. Finally, the large dimension N and time period T result in a vast number of observations, further intensifying the computational burden. Consequently, our methods are designed to rely on the Gibbs sampler whenever possible.

The Markov Chain Monte Carlo (MCMC) algorithm for posterior inference proceeds as follows. Each step is performed conditionally on all other parameters. For clarity, subscripts are omitted, and we present the key elements below; full details are provided in the supplementary material.

1. $\gamma_i \sim N(\bar{m}_{\gamma_i}, \bar{h}_{\gamma_i}^{-1})$, where $\bar{h}_{\gamma_i} = h_\gamma + \sigma_q^{-2}(\mathbf{x}_i^*)' \mathbf{x}_i^*$ and $\bar{m}_{\gamma_i} = \bar{h}_{\gamma_i}^{-1}(h_\gamma m_\gamma + \sigma_q^{-2}(\mathbf{x}_i^*)' \mathbf{y}_i^*)$, $\mathbf{y}_i^* = (y_{i1}^*, \dots, y_{iT}^*)'$, $\mathbf{x}_i^* = (x_{i1}^*, \dots, x_{iT}^*)'$, $y_{it}^* = Q_{it} - [\rho_i \sum_{j=1, j \neq i}^N w_{ij} Q_{jt} + \delta_i \sum_{j=1, j \neq i}^N w_{ij} Q_{j,t-1} + \mathbf{x}'_{it} \mathbf{b}_i + \mathbf{f}'_t \boldsymbol{\lambda}_i]$ and $x_{it}^* = Q_{i,t-1}$.
2. $\delta_i \sim N(\bar{m}_{\delta_i}, \bar{h}_{\delta_i}^{-1})$, where $\bar{h}_{\delta_i} = h_\delta + \sigma_q^{-2}(\mathbf{x}_i^*)' \mathbf{x}_i^*$ and $\bar{m}_{\delta_i} = \bar{h}_{\delta_i}^{-1}(h_\delta m_\delta + \sigma_q^{-2}(\mathbf{x}_i^*)' \mathbf{y}_i^*)$, \mathbf{y}_i^* and \mathbf{x}_i^* are the stacking of y_{it}^* and x_{it}^* , respectively, with $y_{it}^* = Q_{it} - [\rho_i \sum_{j=1, j \neq i}^N w_{ij} Q_{jt} + \gamma_i Q_{i,t-1} + \mathbf{x}'_{it} \mathbf{b}_i + \mathbf{f}'_t \boldsymbol{\lambda}_i]$ and $x_{it}^* = \sum_{j=1, j \neq i}^N w_{ij} Q_{j,t-1}$.
3. $\mathbf{b}_i \sim N(\bar{\mathbf{m}}_{b_i}, \bar{H}_{b_i}^{-1})$, where $\bar{H}_{b_i} = H_b + \sigma_q^{-2}(X_i^*)' X_i^*$ and $\bar{\mathbf{m}}_{b_i} = \bar{H}_{b_i}^{-1}(H_b \mathbf{m}_b + \sigma_q^{-2}(X_i^*)' \mathbf{y}_i^*)$. The vectors \mathbf{y}_i^* and matrix X_i^* are the stacking of y_{it}^* and x'_{it} , respectively with $y_{it}^* = Q_{it} - [\rho_i \sum_{j=1, j \neq i}^N w_{ij} Q_{jt} + \gamma_i Q_{i,t-1} + \delta_i \sum_{j=1, j \neq i}^N w_{ij} Q_{j,t-1} + \mathbf{f}'_{t,\tau} \boldsymbol{\lambda}_{i,\tau}]$ while x_{it} is the data.
4. ρ_i does not have a conjugate representation, ρ_i is a structural parameter from a high-dimensional simultaneous equations system. We propose to draw ρ_i from a two-component mixture normal distribution $\rho_i \sim w_i N(\mu_{1i}, \sigma_{1i}^2) + (1 - w_i) N(\mu_{2i}, \sigma_{2i}^2)$ to approximate the posterior $p(\rho_i | \cdot)$, as being inspired by Villani (2009). The mixture weight w_i and parameters $\mu_{1i}, \mu_{2i}, \sigma_{1i}^2, \sigma_{2i}^2$ are detailed in the Supplement A1.

5. The factor \mathbf{f}_t has a state space representation. The measurement equation is $\mathbf{y}_t^* = \Lambda \mathbf{f}_t + \mathbf{e}_t^q$ with $\mathbf{e}_t^q \sim N(0, \sigma_q^2 I_N)$, where $\mathbf{y}_t^* = (y_{1t}^*, \dots, y_{Nt}^*)$ with $y_{it}^* = Q_{it} - [\rho_i \sum_{j=1, j \neq i}^N w_{ij} Q_{jt} + \gamma_i Q_{i,t-1} + \delta_i \sum_{j=1, j \neq i}^N w_{ij} Q_{j,t-1} + \mathbf{x}'_{it} \mathbf{b}_i]$. Λ is the matrix of λ_{ij} with the top $r \times r$ sub-matrix being lower triangular and positive diagonal elements. The state equation is $\mathbf{f}_t = \Phi \mathbf{f}_{t-1} + \mathbf{u}_t$ with $u_t \sim N(0, I_t)$, where $\Phi = \text{diag}\{\phi_1, \dots, \phi_r\}$. We apply the forward filtering and backward sampling method to draw from its posterior.
6. A truncated normal can make a draw of $\phi_j \mid \cdot \sim N(\bar{m}_{\phi_j}, \bar{h}_{\phi_j}^{-1}) \mathbf{1}(|\phi_j| < 1)$, where $\bar{h}_{\phi_j} = \sum_{t=2}^T f_{j,t-1}^2$ and $\bar{m}_{\phi_j} = (\sum_{t=2}^T f_{j,t} f_{j,t-1}) / (\sum_{t=2}^T f_{j,t-1}^2)$ for $j = 1, \dots, r$.
7. Λ is conditionally Gaussian with simple triangular identification restrictions. For $i \leq r$, denote it as $\beta_i = (\lambda_{i1}, \dots, \lambda_{ii})'$ with $\lambda_{ii} > 0$; otherwise, we randomly draw a vector of length r and denote it as $\beta_i = (\lambda_{i1}, \dots, \lambda_{ir})'$. The conditional distribution of β_i can be summarised in a linear regression as $y_i^* = X_i^* \beta_i + e_i^q$, where y_i^* is the collection of $y_{it}^* = Q_{it} - [\rho_i \sum_{j=1, j \neq i}^N w_{ij} Q_{jt} + \gamma_i Q_{i,t-1} + \delta_i \sum_{j=1, j \neq i}^N w_{ij} Q_{j,t-1} + \mathbf{x}'_{it} \mathbf{b}_i]$. If $i \leq r$, X_i^* is the first i columns from the matrix of factors F , otherwise, X_i^* is simply the full F matrix. We draw β_i from $N(\bar{\mathbf{m}}_{\beta_i}, \bar{H}_{\beta_i}^{-1})$, where $\bar{H}_{\beta_i} = H_{\beta_i} + \sigma_q^{-2} (X_i^*)' X_i^*$ and $\bar{\mathbf{m}}_{\beta_i} = \bar{H}_{\beta_i}^{-1} (\sigma_q^{-2} X_i^*)' \mathbf{y}_i^*$. The H_{β_i} is implied by the prior of Λ .
8. We apply a Metropolis-Hastings algorithm to sample σ based on $y_{it} = Q_{it} + \varepsilon_{it}$, where $\varepsilon_{it} \sim ALD(\tau, \sigma)$.
9. We sample V_{it} by $V_{it} = x^{-1}$ with $x \sim IG(\mu, \lambda)$, where IG means inverse Gaussian with $\mu = \sqrt{\frac{(\xi_1^2 + 2\xi_2^2)\sigma^2}{(y^*)_{it}^2}}$, $\lambda = \frac{\xi_1^2 + 2\xi_2^2}{\xi_2^2}$ with $\xi_1 = \frac{1-2\tau}{\tau(1-\tau)}$, $\xi_2 = \sqrt{\frac{2}{\tau(1-\tau)}}$.
10. The \mathbf{Q}_t has a state space representation. The state equation is $\mathbf{Q}_t = B\mathbf{Q}_{t-1} + X_t + \mathbf{u}_t^q$ with $\mathbf{u}_t^q \sim N(0, \Sigma)$, where $A = (I - \rho W)^{-1}$, $B = (I - \rho W)^{-1}(\gamma + \delta W)$, $X_t = (I - \rho W)^{-1}([xb]_t + \Lambda \mathbf{f}_t)$ and $\Sigma = \sigma_q^2 A A'$. The measurement equation is $\mathbf{Y}_t = \mathbf{Q}_t + \mathbf{u}_t + \mathbf{u}_t^y$, where \mathbf{u}_t is the vectorisation of $\xi_1 u_{it} \equiv \xi_1 \sigma V_{it}$ and $\mathbf{u}_t^y \sim N(0, D_t)$ with $D_t = \text{diag}(\xi_2^2 \sigma^2 V_{1t}, \dots, \xi_2^2 \sigma^2 V_{Nt})$. We apply the forward-filtering and back-ward sampling method.

Proposition 1 *The proposed MCMC sampling procedure, using equation (9), can generate posterior samples from the distribution $\pi(\vartheta_\tau | X, Y)$ in (7) through an importance sampling procedure.*

This section obtains the “pseudo” posterior $p(\vartheta_\tau | X, Y)$. The exact posterior $\pi(\vartheta_\tau | X, Y)$ in (7) can be built through importance sampling. Our simulation study demonstrates that this computationally efficient MCMC approach yields satisfactory results. For example, consider to compute the posterior expectation of $q(\vartheta_\tau)$ with $q(\cdot)$ being a known function of ϑ :

$$E [q(\vartheta_\tau) | X, Y] = \int q(\vartheta_\tau) \pi(\vartheta_\tau | X, Y) d\vartheta_\tau.$$

We can use the importance sampling:

$$E [q(\vartheta_\tau) | X, Y] = \frac{\int q(\vartheta_\tau) \frac{\pi_K(\vartheta_\tau | X, Y)}{p_K(\vartheta_\tau | X, Y)} p(\vartheta_\tau | X, Y) d\vartheta_\tau}{\int \frac{\pi_K(\vartheta_\tau | X, Y)}{p_K(\vartheta_\tau | X, Y)} p(\vartheta_\tau | X, Y) d\vartheta_\tau},$$

where $\pi_K(\vartheta_\tau | X, Y)$ and $p_K(\vartheta_\tau | X, Y)$ are the kernels of $\pi(\vartheta_\tau | X, Y)$ and $p(\vartheta_\tau | X, Y)$, respectively. Because we have a set of posterior samples from $p(\vartheta_\tau | X, Y)$ and the kernels are easy to compute, the posterior expectation of $q(\vartheta_\tau)$ can be obtained without having $\pi(\vartheta_\tau | X, Y)$ directly. Appendix E.3 in supplementary document demonstrates Proposition 1 through a small Monte Carlo simulation.

Note that computing the posterior kernel $p(\vartheta_\tau | X, Y)$ can be computationally expensive. However, after obtaining a sample using our method, it remains feasible, as the importance sampling process is parallelisable. A key aspect of successful importance sampling is having a good proposal distribution, and our method plays a critical role in ensuring that the posterior distribution is precisely targeted.

Remark 3 *This section provides a sufficient approach, particularly when σ_q^2 is tuned for robustness. An alternative way to address the computational burden in the importance sampling is to let $\sigma_q^2 \rightarrow 0$ as the number of MCMC sampling increases. Then, we can have the posterior converging to the true posterior density $\pi(\vartheta_\tau | X, Y)$. As a result, after a certain number of MCMC iterations, the generated sample from $p(\vartheta_\tau | X, Y)$ is regarded as the sample from the true posterior density $\pi(\vartheta_\tau | X, Y)$.*

3.3 Number of Factors

In Bayesian theory, the number of factors can be viewed as a random variable. Its posterior can be inferred from by exploring the marginal likelihood of models with different value of r . However, because of the high computational cost, we adopt the idea of the sparse mixture approach if [Malsiner-Walli et al. \(2016\)](#) by setting a large but finite number of factors. Each factor has a “switch” variables taking value of 0 and 1 indicating whether the corresponding factor is selected. This construction follows the literature of stochastic search variable selection dated back to [Mitchell and Beauchamp \(1988\)](#).

In particular, we revise (9) to have

$$Q_{it} = \rho_{i,\tau} \sum_{j=1, j \neq i}^N w_{ij} Q_{jt} + \gamma_{i,\tau} Q_{i,t-1} + \delta_{i,\tau} \sum_{j=1, j \neq i}^N w_{ij} Q_{j,t-1} + \mathbf{x}'_{it} \mathbf{b}_{i,\tau} + (\mathbf{s}_\tau \circ \mathbf{f}_{t,\tau})' \boldsymbol{\lambda}_{i,\tau} + e_{it,\tau}^q, \quad (12)$$

where \mathbf{s}_τ is a $r_{\max} \times 1$ vector of 0's and 1's and the symbol \circ means the Hadamard product. The maximum number of factors is r_{\max} . Each element $s_{j,\tau}$ in vector \mathbf{s}_τ controls for whether factor $f_{j,\tau}$ is selected.

Define the prior of $\mathbf{s}_\tau = (s_{1,\tau}, \dots, s_{r_{\max},\tau})$ as

$$P(s_{j,\tau} = 1) = \pi, \quad P(s_{j,\tau} = 0) = 1 - \pi$$

for $j = 1, \dots, r_{\max}$. The posterior distribution of the number of factors is the distribution of the sum of \mathbf{s}_τ .

Because the only difference between (9) and (12) is \mathbf{s} , we only need to add one more step for $\mathbf{s}_{u_{it}}$ and revise the step of F slightly in the original MCMC algorithm to infer the number of factors as follows.

1. Draw $s_{k,\tau}$ from a simple Bernoulli distribution. Define

$$y_{it}^* = Q_{it} - \left[\rho_{i,\tau} \sum_{j=1, j \neq i}^N w_{ij} Q_{jt} + \gamma_{i,\tau} Q_{i,t-1} + \delta_{i,\tau} \sum_{j=1, j \neq i}^N w_{ij} Q_{j,t-1} + \mathbf{x}'_{it} \mathbf{b}_{i,\tau} + (\mathbf{s}_\tau \circ \mathbf{f}_{t,\tau})' \boldsymbol{\lambda}_{i,\tau} \right]$$

$$p(s_{k,\tau} = m \mid \cdot) \propto p(s_{k,\tau} = m) \prod_{i=1}^N \prod_{t=1}^T f_N(y_{it}^* \mid 0, \sigma_q^2),$$

where $m = 0$ or 1 , with the corresponding $s_{k,\tau}$ being set to 0 or 1 . There is a slight abuse of notation. Namely, y_{it}^* is different when $s_{k,\tau} = 0$ and 1 .

2. To draw $f_{t,\tau}$, we split this into two parts.
 - (a) Only draw the active factors for $s_{k,\tau} = 1$.
 - (b) Conditional on the active factors, draw the other factors similar as in the MCMC step. This approach follows the Reversible jump MCMC method in [Green \(1995\)](#).

Remark 4 *In practice, we tune σ_q^2 to control an R^2 level from (9) or (12). Alternatively, if we do not tune but want estimate σ_q^2 we can monitor the R^2 instead. For a high R^2 (in our application more than 98%), the parameters can be used as if they were drawn from the true posterior of the model. Alternatively, one can use the posterior sample from this model setting as a proposal distribution for an importance sampling scheme applied to the original model. Because the importance sampling is parallelisable, such second stage computation is much more affordable than any generic methods.*

4 Asymptotic results

To provide a theoretical justification for the Bayesian method, this section presents results on posterior consistency. A sequence of posterior distributions is considered consistent if, as the length of the time series and the number of cross-sectional units increase, the posterior converges to the degenerate measure at the true parameter value of the population density. Intuitively, posterior consistency ensures that the information from the quantile objective function outweighs the prior information. Before analysing the asymptotic behavior of the posterior distribution, we first need to examine the average consistency of the frequentist estimator. This is necessary to establish a set of proper conditions (conditions that have not yet been fully explored) that guarantee the convergence of the estimated model to the true population density.

4.1 Assumptions

Below, we denote the true spatial parameters, the lag coefficients and the true regression coefficient as $\rho_{i,0,\tau}$, $\delta_{i,0,\tau}$, $\gamma_{i,0,\tau}$ and $\mathbf{b}_{i,0,\tau}$, respectively. Similarly, we denote $F_{0,\tau} = (\mathbf{f}_{1,0,\tau}, \dots, \mathbf{f}_{T,0,\tau})'$ and $\Lambda_{0,\tau} = (\boldsymbol{\lambda}_{1,0,\tau}, \dots, \boldsymbol{\lambda}_{N,0,\tau})'$ as the true factors and loadings. A set of regularity conditions that are needed for theoretical analysis are given as follows.

Assumption A: Common factors

Let \mathcal{F} be a compact subset of \mathbb{R}^{r_τ} . The common factors $\mathbf{f}_{t,0,\tau} \in \mathcal{F}$ satisfy $T^{-1} \sum_{t=1}^T \mathbf{f}_{t,0,\tau} \mathbf{f}'_{t,0,\tau} = I_{r_\tau}$.

Assumption B: Factor loadings, the lag coefficients and regression coefficients

- (B1) Let $\mathcal{P}, \mathcal{D}, \mathcal{G}$ be compact subsets of \mathbb{R} , and let \mathcal{B} and \mathcal{L} be compact subsets of \mathbb{R}^{p+1} and \mathbb{R}^{r_τ} , respectively. The spatial parameters $\rho_{i,0,\tau}$ and $\delta_{i,0,\tau}$, the lag coefficients $\gamma_{i,0,\tau}$, the regression coefficient $\mathbf{b}_{i,0,\tau}$, and the factor-loading $\boldsymbol{\lambda}_{i,0,\tau}$ satisfy that $\rho_{i,0,\tau} \in \mathcal{P}$, $\delta_{i,0,\tau} \in \mathcal{D}$, $\gamma_{i,0,\tau} \in \mathcal{G}$, $\mathbf{b}_{i,0,\tau} \in \mathcal{B}$ and $\boldsymbol{\lambda}_{i,0,\tau} \in \mathcal{L}$ for each i .
- (B2) The factor-loading matrix $\Lambda_{0,\tau} = (\boldsymbol{\lambda}_{1,0,\tau}, \dots, \boldsymbol{\lambda}_{N,0,\tau})'$ satisfies $N^{-1} \sum_{i=1}^N \boldsymbol{\lambda}_{i,0,\tau} \boldsymbol{\lambda}'_{i,0,\tau} \xrightarrow{p} \Sigma_{\Lambda_\tau}$, where Σ_{Λ_τ} is an $r_\tau \times r_\tau$ positive definite diagonal matrix with diagonal elements distinct and arranged in the descending order. In addition, the eigenvalues of Σ_{Λ_τ} are distinct.

Assumption C: Idiosyncratic error terms

(C1): The random variable

$$\varepsilon_{it,\tau} = y_{it} - Q_{y_{it}} \left(\tau | X_t, B_\tau, \mathbf{f}_{t,\tau}, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\delta}_\tau, \boldsymbol{\gamma}_\tau \right)$$

satisfies $P(\varepsilon_{it,\tau} \leq 0) = \tau$, and is independently distributed over i and t , conditional on $X_t, B_{0,\tau}, F_{0,\tau}, \Lambda_{0,\tau}, \boldsymbol{\rho}_{0,\tau}, \boldsymbol{\delta}_{0,\tau}$ and $\boldsymbol{\gamma}_{0,\tau}$.

(C2): The conditional density function of $\varepsilon_{it,\tau}$ given $\{X_t, B_{0,\tau}, F_{0,\tau}, \Lambda_{0,\tau}, \boldsymbol{\rho}_{0,\tau}, \boldsymbol{\delta}_{0,\tau}, \boldsymbol{\gamma}_{0,\tau}\}$, denoted as $g_{it}(\varepsilon_{it,\tau})$, is continuous. In addition, for any compact set \mathcal{C} , there exists a positive constant $\underline{g} > 0$ (depending on \mathcal{C}) such that $\inf_{c \in \mathcal{C}} g_{it}(c) \geq \underline{g}$ for all i and t .

Assumption D: Weight matrix

(D1): W is an exogenous spatial weights matrix whose diagonal elements of W are all zeros. In addition, W is bounded by some constant C for all N under $\|\cdot\|_1$ and $\|\cdot\|_\infty$.

(D2): The matrix $(I_N - \rho_\tau W)^{-1}$ satisfies $\sup_{\rho_\tau \in \mathcal{P}} \|I_N - \rho_\tau W\|_2 < C$, and

$$\sup_{\rho_\tau \in \mathcal{P}} \left(\left\| (I_N - \rho_\tau W)^{-1} \right\|_1 \vee \left\| (I_N - \rho_\tau W)^{-1} \right\|_\infty \right) < C$$

where C is some positive constant.

Assumption E: Explanatory variables and design matrix

(E1): For a positive constant C , explanatory variables satisfy $\sup_{it} \|\mathbf{x}_{it}\| \leq C$ almost surely.

(E2): Denote $\mathbf{Q}_{t,\tau}^0 = \mathbf{Q}_t(X_t, F_{t,0,\tau}, B_{0,\tau}, \Lambda_{0,\tau}, \rho_{0,\tau}, \gamma_{0,\tau}, \delta_{0,\tau})$. Let $\mathcal{X}(B_{0,\tau})$ be an $N \times T$ matrix with its (i, t) -th entry $\mathbf{x}'_{it} \mathbf{b}_{0,\tau}$. Define $u_{it,0,\tau}$ to be the (i, t) th element of $\mathcal{U}_{0,\tau}$ with

$$\mathcal{U}_{0,\tau} = W(I_N - \rho_{0,\tau} W)^{-1} \left[\mathcal{X}(B_{0,\tau}) + \Lambda_{0,\tau} F'_{0,\tau} + (\gamma_{0,\tau} + \delta_{0,\tau} W) \mathcal{Q}_{-1,\tau}^0 \right].$$

where $\mathcal{Q}_{-1,\tau}^0 = [\mathbf{Q}_{0,\tau}^0, \mathbf{Q}_{1,\tau}^0, \dots, \mathbf{Q}_{T-1,\tau}^0]$. Let $\mathbf{z}_{it,\tau} = (u_{it,0,\tau}, Q_{i,t-1,\tau}^0, \sum_{j=1}^N w_{ij} Q_{j,t-1,\tau}^0, \mathbf{x}'_{it})'$ with $Q_{i,t-1,\tau}^0$ being the i -th element of $\mathbf{Q}_{t-1,\tau}^0$, and $\mathbf{Z}_{i,\tau} = (\mathbf{z}_{i1,\tau}, \mathbf{z}_{i2,\tau}, \dots, \mathbf{z}_{iT,\tau})'$. Further define $A_{i,\tau} = \frac{1}{T} \mathbf{Z}'_{i,\tau} M_{F_\tau} \mathbf{Z}_{i,\tau}$, $B_{i,\tau} = (\lambda_{i,0,\tau} \boldsymbol{\lambda}'_{i,0,\tau}) \otimes I_T$, $C_{i,\tau} = \frac{1}{\sqrt{T}} [\boldsymbol{\lambda}_{i,0,\tau} \otimes (M_{F_\tau} \mathbf{Z}_{i,\tau})]'$ with $M_{F_\tau} = I - F_\tau (F'_\tau F_\tau)^{-1} F'_\tau$. Let \mathcal{F}_τ be the collection of F_τ such that $\mathcal{F}_\tau = \{F_\tau : F'_\tau F_\tau / T = I_{r_\tau}\}$. We assume that with probability approaching one,

$$\inf_{F_\tau \in \mathcal{F}_\tau} \lambda_{\min} \left[\frac{1}{N} \sum_{i=1}^N E_{i,\tau}(F_\tau) \right] > 0,$$

where $\lambda_{\min}(A)$ denotes the smallest eigenvalue of matrix A , and $E_{i,\tau}(F_\tau) = B_{i,\tau} - C'_{i,\tau} A^{-1}_{i,\tau} C_{i,\tau}$.

(E3): For each i , we assume that there exists a constant $c > 0$ such that for each i , with probability approaching one,

$$\liminf_{T \rightarrow \infty} \lambda_{\min} \left(\frac{1}{T} \mathbf{Z}'_{i,\tau} M_{F_{0,\tau}} \mathbf{Z}_{i,\tau} \right) \geq c.$$

Assumption F: Stationary condition

The data generating process from (1) is assumed to be stationary. To ensure the stationarity, it is assumed that $\bar{c} < 1$ where

$$\bar{c} = \sup_{\rho_\tau \in \mathcal{P}, \delta_\tau \in \mathcal{D}, \gamma_\tau \in \mathcal{G}} \left| \lambda_{\max} \left(A(\rho_\tau, \delta_\tau, \gamma_\tau) \right) \right| < 1.$$

where $\lambda_{\max}(A)$ denotes the eigenvalue of A with the largest modulus.

Remark 5 *Assumptions A and B on the factors and factor loadings are standard in factor models. Similar to Ando and Bai (2020), the factors and factor loadings are treated as parameters. Assumptions C and D on the idiosyncratic errors and the spatial weighting matrix are standard assumptions in the literature. Assumption E is necessary for deriving the consistency of the frequentist estimator (See Ando and Bai (2020) for similar assumptions). Assumption F is a stationary condition similar to Yu et al. (2008). Similar to the investigation in Yu et al. (2008), a sufficient condition for Assumption F is $\sup \|A(\rho_\tau, \delta_\tau, \gamma_\tau)\|_2 < 1$. The stationary condition is verified accordingly in both simulation and empirical analysis.*

Now, we investigate the consistency of the frequentist estimator, defined as the minimiser of (6) subject to a normalisation condition. Recall $\vartheta_\tau = \{\rho_\tau, \delta_\tau, \gamma_\tau, B_\tau, \Lambda_\tau, F_\tau\}$, and $\hat{\vartheta}_\tau = \{\hat{\rho}_\tau, \hat{\delta}_\tau, \hat{\gamma}_\tau, \hat{B}_\tau, \hat{\Lambda}_\tau, \hat{F}_\tau\}$ denotes the frequentist estimator. A set of assumption A–F leads to the following result.

Theorem 1 *Suppose that the number of common factors in (1) is correctly specified. Under Assumptions A–F, $\log(N)/T \rightarrow 0$ as $N, T \rightarrow \infty$, the frequentist estimator is the consistent estimator for their true values in the sense that*

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \|\hat{\rho}_{i,\tau} - \rho_{i,\tau,0}\|^2 &= O_p(\delta_{NT}^2), & \frac{1}{N} \sum_{i=1}^N \|\hat{\gamma}_{i,\tau} - \gamma_{i,\tau,0}\|^2 &= O_p(\delta_{NT}^2), \\ \frac{1}{N} \sum_{i=1}^N \|\hat{\delta}_{i,\tau} - \delta_{i,\tau,0}\|^2 &= O_p(\delta_{NT}^2), & \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{b}}_{i,\tau} - \mathbf{b}_{i,\tau,0}\|^2 &= O_p(\delta_{NT}^2), \\ \frac{1}{NT} \|\hat{\Lambda}_\tau \hat{F}'_\tau - \Lambda_{\tau,0} F'_{\tau,0}\|^2 &= O_p(\delta_{NT}^2). \end{aligned} \tag{13}$$

where $\delta_{NT} = \max(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{T}})$.

Remark 6 The last claim $\frac{1}{NT} \|\hat{\Lambda}_\tau \hat{F}'_\tau - \Lambda_{\tau,0} F'_{\tau,0}\|^2 = O_p(\delta_{NT}^2)$ further implies

$$\frac{1}{N} \sum_{i=1}^N \|\hat{\lambda}_{i,\tau} - \lambda_{i,\tau,0}\|^2 = O_p(\delta_{NT}^2), \quad \frac{1}{T} \sum_{t=1}^T \|\hat{\mathbf{f}}_{t,\tau} - \mathbf{f}_{t,\tau,0}\|^2 = O_p(\delta_{NT}^2).$$

Remark 7 The structure (4) indicates that the quintile function at time t is expressed as the sum of the previous quintile functions up to time $t - 1$. This created a difficult technical challenge to establish the claims in Theorem 1. More specifically, as an intermediate result, the technical proof of our Lemma 2 (in the supplementary document) establishes $\frac{1}{N} \sum_{i=1}^N \|\hat{\phi}_{i,\tau} - \phi_{i,\tau,0}\|^2 = o_p(1)$, where $\phi_{i,\tau,0} = (\rho_{i,\tau,0}, \gamma_{i,\tau,0}, \delta_{i,\tau,0}, \mathbf{b}'_{i,\tau,0})'$, $\hat{\phi}_{i,\tau} = (\hat{\rho}_{i,\tau}, \hat{\gamma}_{i,\tau}, \hat{\delta}_{i,\tau}, \hat{\mathbf{b}}'_{i,\tau})'$. This result was not available from previous studies (Ando and Bai (2020), Chen et al. (2021)) due to the static nature of their model. This illustrates one of the key challenges when we establish Theorem 1.

The next theorem also plays an important role when we investigate the consistency of our Bayesian MCMC procedure. Theorem 2 implies that it is ideal to set the number of common factors equal to or greater than the true number of common factors when one's focus is the consistent estimation of parameters ρ_τ , δ_τ , γ_τ and B_τ . To obtain the claim, we need additional assumption.

Assumption G: Identification of B_τ for over-fitted model

Let $F_\tau(k)$ be the common factor matrix with $k > r$ and r being the true number of common factors, $\mathcal{Z}_\tau(\phi_\tau)$ be the $N \times T$ matrix with its (i, t) th entry equal to $z'_{it,\tau} \phi_{i,\tau}$, where $\phi_\tau = (\phi_{1,\tau}, \phi_{2,\tau}, \dots, \phi_{N,\tau})'$ and $\phi_{i,\tau} = (\rho_{i,\tau}, \gamma_{i,\tau}, \delta_{i,\tau}, \mathbf{b}'_{i,\tau})'$. Here $z_{it,\tau}$ is defined in Assumption E.2. There exists a positive constant $\check{c} > 0$ such that with probability approaching one,

$$\inf_{F_\tau(k), F_\tau(k)' F_\tau(k) / T = I_k} \frac{1}{NT} \|M_{\Lambda_{0,\tau}} \mathcal{Z}_\tau(\phi_\tau) M_{F_\tau(k)}\|^2 \geq \check{c} \frac{1}{N} \sum_{i=1}^N \|\phi_{i,\tau}\|^2,$$

where $M_{\Lambda_{0,\tau}} = I - \Lambda_{0,\tau} (\Lambda'_{0,\tau} \Lambda_{0,\tau})^{-1} \Lambda'_{0,\tau}$.

Theorem 2 Suppose that the specified number of common factors in (1) is larger than the true number of common factors. Under Assumptions A–G, $\log(N)/T \rightarrow 0$ as $N, T \rightarrow \infty$, the corresponding frequentist estimator still satisfies the claims in (13).

Now, our concern is the sequence of posterior distributions $\pi(\vartheta_\tau|Y, X)$ constructed by the size of $T \times N$ panel data, generated from the true density $f(Y|X, \vartheta_{\tau,0})$. In this paper, we show that the constructed posterior forms a Hellinger-consistent sequence. In regards to the posterior consistency based on Hellinger distance, we refer to [Barron et al. \(1999\)](#), [Ghosal et al. \(1999\)](#), [Walker and Hjort \(2001\)](#).

Recall the pseudo-likelihood based density function:

$$f(Y|X, \vartheta_\tau) \propto \exp \left[-\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T q_{it,\tau}(\vartheta_\tau) \right],$$

where $q_{it,\tau}(\vartheta_\tau) \equiv q_\tau(y_{it} - Q_{y_{it}}(\tau|X_t, F_{t,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau))$. We establish Bayesian consistency under the pseudo-likelihood $f(Y|X, \vartheta_\tau)$, in the sense that, for any $\mu > 0$,

$$\pi_{N,T,\tau}(\{\vartheta_\tau : H(f(Y|X, \vartheta_\tau), f(Y|X, \vartheta_{\tau,0})) > \mu\}) \rightarrow 0 \text{ as } N, T \rightarrow \infty, \quad (14)$$

where $\vartheta_{\tau,0}$ is true value of ϑ_τ , $\pi_{N,T,\tau}(\cdot)$ is defined as

$$\pi_{N,T,\tau}(A) = \int_A \frac{f(Y|X, \vartheta_\tau)}{f(Y|X, \vartheta_{\tau,0})} \pi(\vartheta_\tau) d\vartheta_\tau,$$

where $A \subset \Theta$ is the subset of parameter space Θ , and for two density functions $h(y)$ and $g(y)$, the Hellinger distance is defined as $H(h, g) = \left\{ \int (g^{1/2}(y) - h^{1/2}(y)) d(y) \right\}^2$. To obtain the result in (14), we need an additional condition.

Assumption H: Kullback–Leibler property

Let $K_\varepsilon(\vartheta_{\tau,0})$ be a Kullback–Leibler neighborhood of $\vartheta_{\tau,0}$ such that ϑ_τ satisfies

$$\int \log\{f(Y|X, \vartheta_{\tau,0})/f(Y|X, \vartheta_\tau)\} f(Y|X, \vartheta_{\tau,0}) dY < \varepsilon.$$

Then, the prior density $\pi(\vartheta_\tau)$ assigns positive mass on all Kullback–Leibler neighborhoods of the pseudo-likelihood based density under the true value $f(Y|X, \vartheta_{\tau,0})$, $\pi(K_\varepsilon(\vartheta_{\tau,0})) > 0$.

Theorem 3 *Under Assumptions A–H, as N, T go to infinity with $N/T \rightarrow 0$, then result (14) holds.*

When Bayesian consistency is considered important, the above theorem offers guidance for designing an appropriate prior distribution. The prior distribution discussed in Section 3 is specifically constructed to satisfy Assumption H. As a result, we expect that the posterior mean under our prior will converge to the true parameter values as both N and T tend to infinity. This expectation is confirmed through our simulation study.

Remark 8 *A common question is about the asymptotic properties of the posterior distribution regarding the number of common factors in our MCMC procedure. When the number of common factors is strictly smaller than the true number, there exists a positive constant such that the expected quantile loss is larger than that under the true number of common factors. As a result, our MCMC procedure asymptotically eliminates posterior samples with r smaller than the true number of common factors as both N and T tends to infinity.*

5 Analysis of gasoline price

We apply our method to the fuel prices reported by retailers in Queensland, a state located in the northeast of Australia. In Queensland, an aggregation system for fuel price reporting has been established under Section 4 of the Fair Trading (Fuel Price Reporting) Regulation 2018.¹ As a result, all fuel retailers in Queensland (including all fuel stations) are required to report their fuel prices as part of the Queensland fuel price reporting scheme, which helps motorists find the cheapest fuel prices. This requirement has been in effect since 3 December 2018. The data is publicly available at <https://www.data.qld.gov.au/dataset/fuel-price-reporting>.

Figure 1 (a) shows the locations of these stations across Queensland, as well as their brands.² In total, there are 1011 registered stations in Queensland. Figure 1 (b) displays the locations of fuel stations in Brisbane, the capital city of Queensland, along with their respective brands. It is evident that many stations are situated close to one another, particularly in the municipal area.

¹See <https://www.epw.qld.gov.au/about/initiatives/fuel-price-reporting>

²The brand small aggregates the smaller brands with fewer than 5 stations.

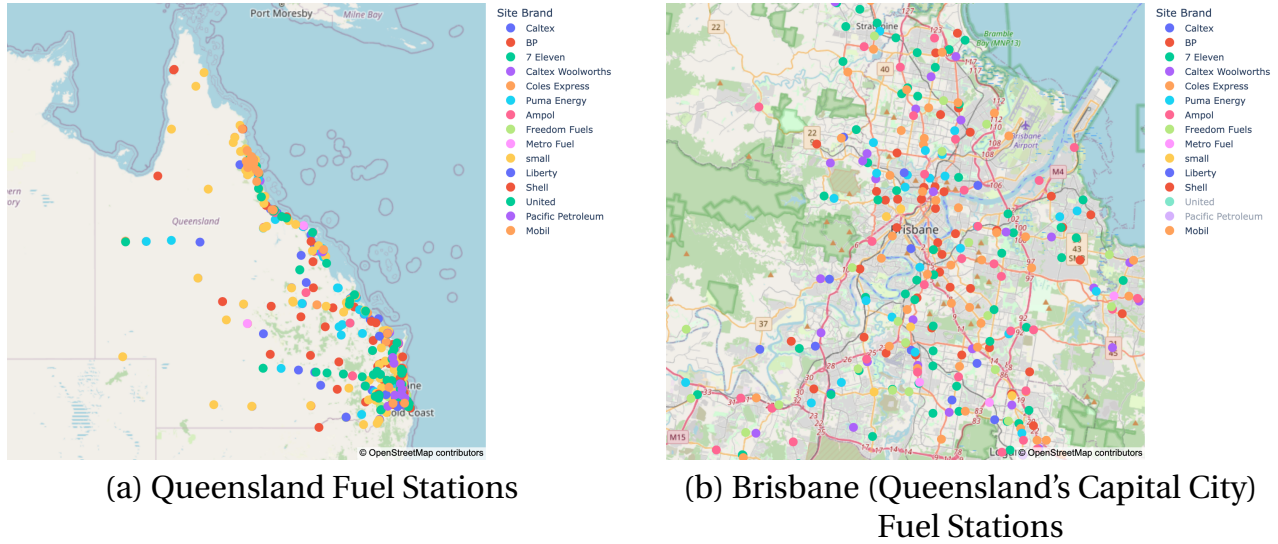


Figure 1: Maps of fuel stations: (a) Queensland, (b) Brisbane.

5.1 Data

We analyse unleaded gasoline, as in [Pinkse et al. \(2002\)](#), because it has the largest market share and is nearly a homogeneous product. Our data spans from February 1, 2019, to September 30, 2021, with a total of $T = 973$ days and no missing values. This sample includes five lockdown periods in Queensland during the Covid-19 pandemic.

Figure 2 shows the average fuel prices for each brand over time, revealing clear seasonality. The significant drop in fuel prices in 2020 corresponds to the longest lockdown period, from March 26 to the end of April. Additionally, when aggregating over time, Figure 3 highlights the price heterogeneity across brands.

In the following analysis, we exclude small brands with fewer than 5 stations and stations located on islands. After this cleaning process, we are left with $N = 946$ stations. Throughout the sample period, no new stations were built, nor were any stations decommissioned. While some stations changed brands, we account for the brand effect in the estimation.

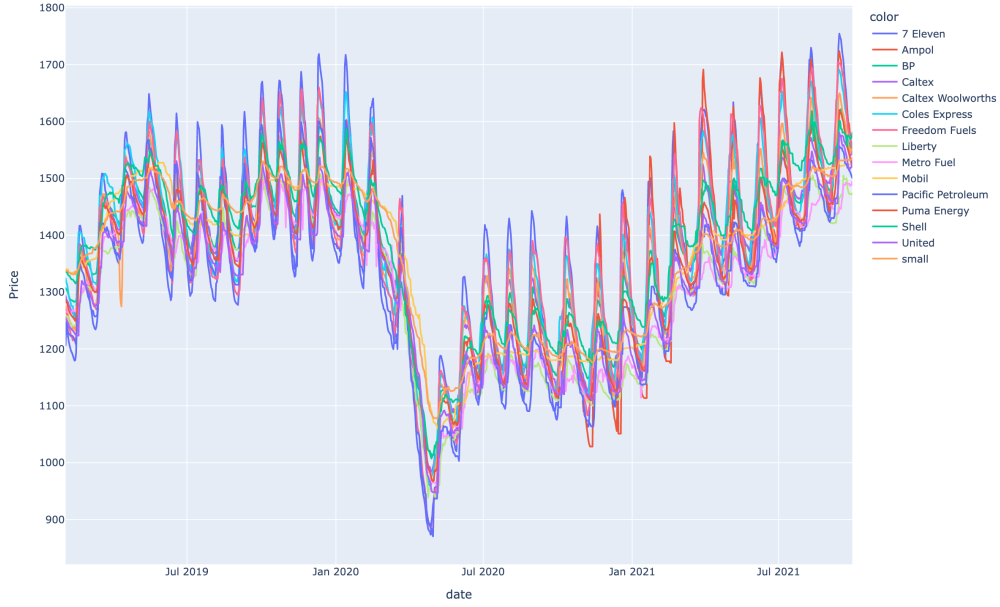


Figure 2: Average price by brands over time
The price unit is 1/10 cent per liter. Namely, 1200 means 1.2 Australian dollar per litre.

5.2 Empirical model specification and estimation

For the term $\mathbf{x}'_{it} \mathbf{b}_{i,\tau}$ in (1), we apply the following empirical specification:

$$\begin{aligned} \mathbf{x}'_{it} \mathbf{b}_{i,\tau} = & \text{Covid}_t \times b_{\text{covid},i,\tau} + \text{Brand}_{it} \times b_{\text{brand}_{it},\tau} + \text{Year}_t \times b_{\text{Year}_{it},\tau} + \text{Month}_t \times b_{\text{Month}_{it},\tau} \\ & + \text{Day of Week}_t \times b_{\text{DayofWeek}_{it},\tau} + G_{i,e_{it}}^{-1}(\tau) \end{aligned} \quad (15)$$

In the above equation, the subscript i refers to fuel station i , and t represents time (day). The independent variables in this model include several time effects: Year_t , Month_t , and DayofWeek_t . The year effect captures any overall trend, while the month and day-of-the-week effects capture explicit seasonal variations. Each station i has its own parameter for these time variables. For instance, the effect of the year 2021 differs for station 1 compared to station 2.

The brand Brand_{it} dummy variables capture the brand effect. Note that some stations changed brands during the sample period, which is why a double subscript is used for this variable. We include all brand dummy variables and exclude the intercept for identification purposes, so the station-specific effect is naturally incorporated.

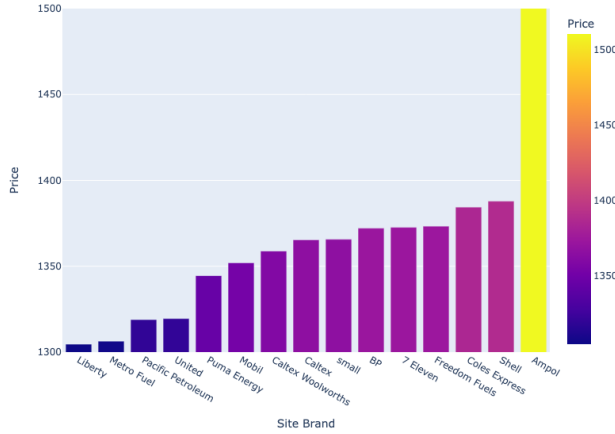


Figure 3: Average price by brands

The variable Covid_t is a dummy variable that takes the value 1 if time t falls within a lockdown period and 0 otherwise. Details of the lockdown periods are provided in the appendix. Briefly, there was only one long lockdown period in 2020, from March 26 to the end of April, with all other periods being less than one week in duration.

For the weight matrix W , we consider **driving** distance rather than geographic distance, taking into account traffic conditions and speed limits in different areas. To compute the average driving time between two stations, we use the Open Source Routing Machine (OSRM). In this application, if two stations are within a 5-minute driving distance of each other, they are classified as neighbors. We normalise each row of the matrix W such that the sum of the elements in each row equals 1. Specifically, if station 1 has three neighbors (stations 2, 7, and 9), then $W_{1,2} = W_{1,7} = W_{1,9} = \frac{1}{3}$, and all other $W_{1,j} = 0$.

The prior is set to be informative but covers a broad range of the parameter space, as in the simulation. The detailed settings can be found in the appendix.

5.3 Results

Due to the large number of parameters in the model, such as the number of ρ , δ , γ , and b , Λ , F , which totals $3N + N(p+1) + r_f(T+N) = 44,597$ in our application, we do not store all simulated values during the MCMC process. Instead, we focus on the posterior means, which allows us to accumulate values with minimal memory usage. This approach makes

it straightforward to evaluate uncertainties. For example, if posterior variance is required, we can save the sum of the squared values, then use the sample mean of the squared values and the sample mean to compute the sample variance. Any moment-based posterior statistics can be derived in this way.

Figure 4 presents the histogram of the posterior means of $\rho_{i,\tau}$ for all i at the quantiles $\tau = (0.01, 0.05, 0.5, 0.95, 0.99)$. Without imposing any restrictions, the distribution of $\rho_{i,\tau}$ reveals two key characteristics. First, the values are positive, indicating that positive spillover effects exist between fuel prices at nearby stations. Second, the heterogeneity patterns are consistent across different quantiles. Some stations exhibit greater sensitivity to their neighbors (larger ρ values), while others show minimal sensitivity, with ρ values close to zero. Figure 4 underscores the need for a heterogeneous coefficient model.

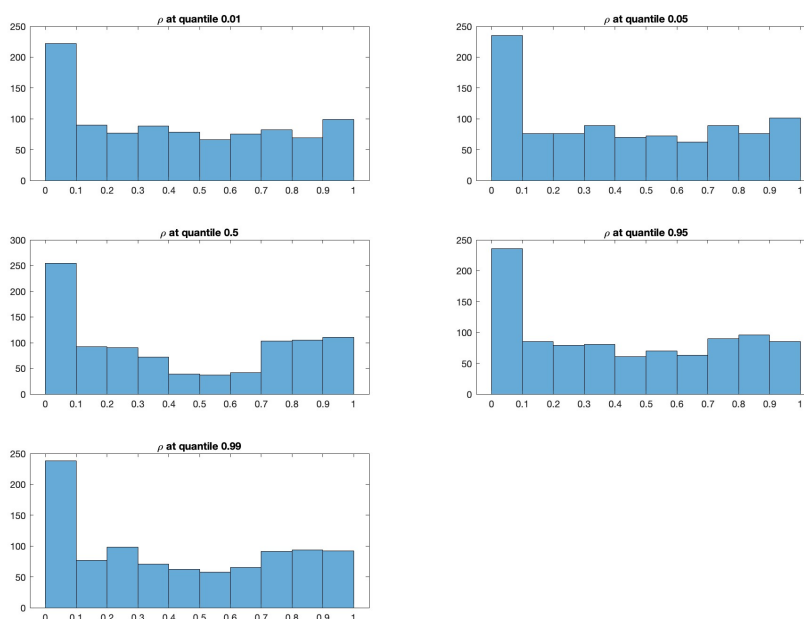


Figure 4: Distributions of ρ

Figure 5 shows the distribution of the posterior means of the Covid lockdown coefficients at different quantiles. It is clear that the distributions of these coefficients differ significantly across quantiles. For instance, the histogram for $\tau = 0.05$ is right-skewed, while the histogram for $\tau = 0.99$ is left-skewed. The mode is around zero. This is not

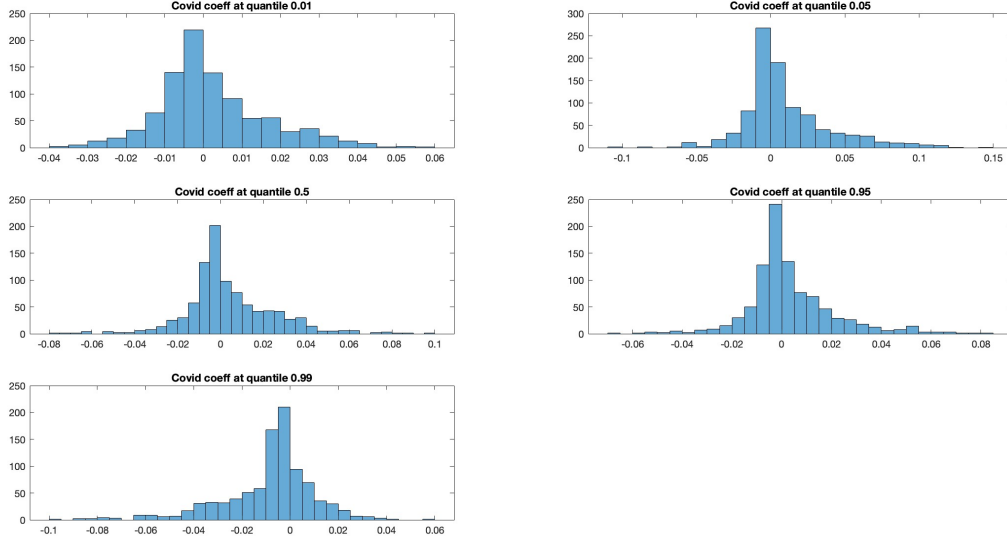


Figure 5: Distribution of Covid Coefficients

surprising, as the factors are intended to capture any systematic price changes. Figure 5 demonstrates that the lockdowns have caused changes in prices, not only in terms of dispersion but also in how these price responses vary across different quantiles.

Figure 6 displays the average brand premiums across different quantiles. It is important to note that each station has its own distinct brand premium, which can be interpreted as the average station effect within the same brand. Similar patterns emerge across the brand premiums. For instance, Pacific Petroleum consistently has the lowest values, while United Petroleum consistently has the highest values across all quantiles.

Similar to the spatial coefficient $\rho_{i,\tau}$, the lag coefficient $\gamma_{i,\tau}$ and the lag-spatial coefficient $\delta_{i,\tau}$ demonstrate a strong pattern of heterogeneity, while exhibiting similar patterns across quantiles, respectively. Due to page limitations, these are not shown here.

The systemic factors and their effects on each time series are plotted in Figure 7. Each subplot represents $N = 946$ time series, depicting $\lambda'_{i,\tau} f_{t,\tau}$. It is evident that the factor structure captures the seasonality present in the data. Since the simulation study does not indicate the correct number of factors, we refrain from analysing individual factors in this application.

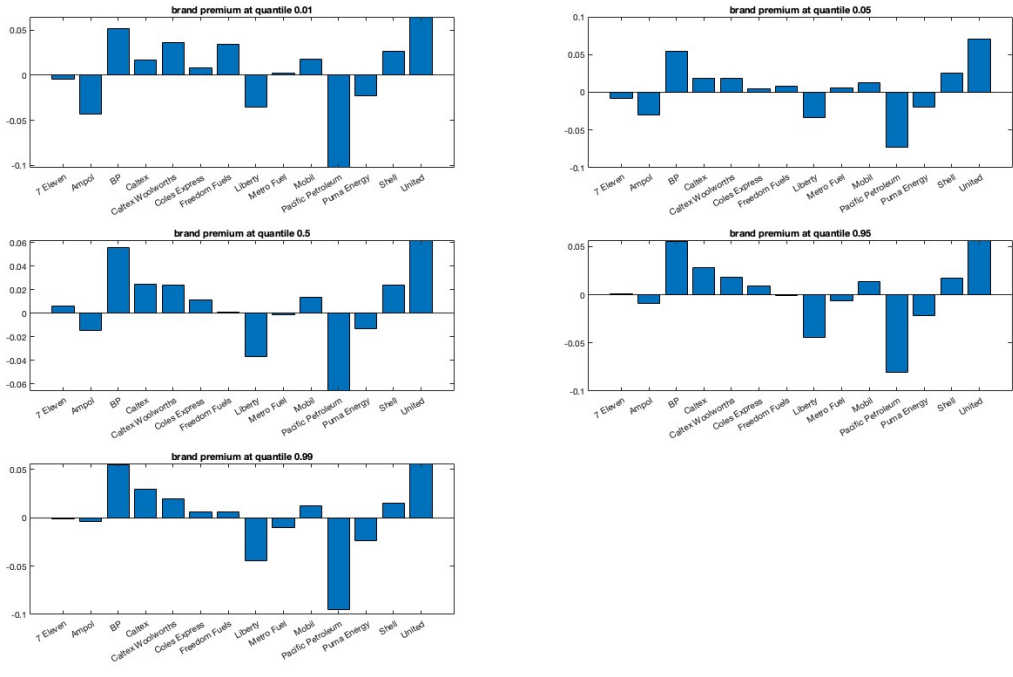


Figure 6: Average Brand Premiums

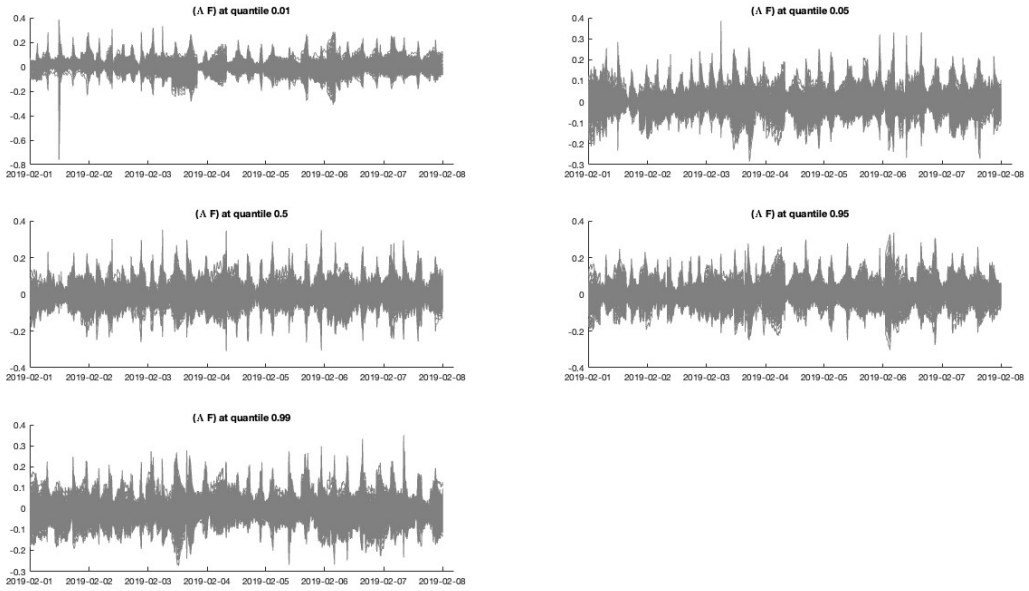


Figure 7: Posterior mean of λF

We compute the average variation from the posterior mean of the contemporaneous effect. Specifically, we consider the term $\rho_{i,\tau} \sum_{i \neq j, j=1}^N w_{ij} Q_{y_{jt}}(\tau | X_t, F_{t,\tau}, B_\tau, \Lambda_\tau, \boldsymbol{\rho}_\tau, \boldsymbol{\gamma}_\tau, \boldsymbol{\delta}_\tau)$ in (1). The ratio of the average variation of this term to the average variation of the posterior values of the quantiles is approximately 30% for all quantiles in (0.01, 0.05, 0.5, 0.95, 0.99). This suggests that the contemporaneous spatial effect plays a significant role in explaining the quantiles.

6 Conclusion

In this paper, we proposed a novel dynamic spatial panel quantile model with interactive effects. The model effectively captures several complex features simultaneously, including spatial spillover effects, heterogeneous regression coefficients, and unobserved heterogeneity that vary across quantiles. To estimate the model, we developed a Bayesian MCMC procedure capable of handling all these aspects, and we established Bayesian consistency to support its theoretical validity. The practical utility of the method was demonstrated through an application to gasoline price data in Australia.

Supplementary Materials

All technical proofs of theoretical results and some numerical results are delegated to the supplementary document.

References

- Ando, T. and J. Bai (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. Journal of the American Statistical Association 112(519), 1182–1198.
- Ando, T. and J. Bai (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. Journal of the American Statistical Association 115(529), 266–279.

- Anselin, L. (1988). *Spatial econometrics: methods and models*.
- Aquaro, M., N. Bailey, and M. H. Pesaran (2021). Estimation and inference for spatial models with heterogeneous coefficients: an application to us house prices. *Journal of Applied Econometrics* 36(1), 18–44.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77(4), 1229–1279.
- Bai, J. and K. Li (2013). Spatial panel data models with common shocks. manuscript.
- Bai, J. and K. Li (2014). Theory and methods of panel data models with interactive effects. *The Annals of Statistics* 42(1), 142–170.
- Bai, J. and K. Li (2021). Dynamic spatial panel data models with common shocks. *Journal of Econometrics* 224(1), 134–160.
- Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of econometrics* 176(1), 18–29.
- Baltagi, B. (2011). *Spatial Panels*, Chapter 15, pp. 435–454. Chapman and Hall.
- Barron, A., M. J. Schervish, and L. Wasserman (1999). The consistency of posterior distributions in nonparametric problems. *Annals of Statistics* 27, 536–561.
- Beenstock, M. and D. Felsenstein (2015). Estimating spatial spillover in housing construction with nonstationary panel data. *Journal of Housing Economics* 28, 42–58.
- Chen, L., J. Gonzalo, and J. Dolado (2021). Quantile factor models. *Econometrica* 89, 875–910.
- Ghosal, S., J. K. Ghosh, and R. V. Ramamoorthi (1999). Posterior consistency of dirichlet mixtures in density estimation. *Annals of Statistics* 27, 143–158.
- Glaser, S., R. Jung, and K. Schweikert (2022). Spatial panel count data: modeling and forecasting of urban crimes. *Journal of Spatial Econometrics* 3(2).

- Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. Biometrika 82(4), 711–732.
- Hallin, M. and R. Liška (2007). The generalized dynamic factor model: determining the number of factors. Journal of the American Statistical Association 102, 603–617.
- Hunneman, A., J. Elhorst, and T. Bijmolt (2022). Store sales evaluation and prediction using spatial panel data models of sales components. Spatial Economic Analysis 17, 127–150.
- Kelejian, H. H. and I. R. Prucha (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. Journal of Econometrics 118, 27–50.
- Koenker, R. and G. Bassett (1978). Regression quantiles. Econometrica 46, 33–50.
- Koenker, R. and Z. Xiao (2006). Quantile autoregression. Journal of the American Statistical Association 101(9), 980–990.
- Kozumi, H. and G. Kobayashi (2011). Gibbs sampling methods for bayesian quantile regression. Journal of statistical computation and simulation 81(11), 1565–1578.
- Lee, L. (2004). Asymptotic distributions of quasi-maximum likelihood estimator for spatial autoregressive models. Econometrica 72(6), 1899–1925.
- Li, K. (2017). Fixed-effects dynamic spatial panel data models and impulse response analysis. Journal of Econometrics 198(1), 102–121.
- Lin, X. and L. Lee (2010). Gmm estimation of spatial autoregressive models with unknown heteroscedasticity. Journal of Econometrics 157, 34–52.
- Lu, L. (2017). Simultaneous spatial panel data models with common shocks. RPA working paper RPA17-03, Federal Reserve Bank of Boston.
- Lu, X. and L. Su (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. Journal of Econometrics 190(1), 148–175.
- Malsiner-Walli, G., S. Frühwirth-Schnatter, and B. Grün (2016). Model-based clustering based on sparse finite gaussian mixtures. Statistics and computing 26(1-2), 303–324.

- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. Review of Economic Studies 60.
- Mitchell, T. J. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. Journal of the American Statistical Association 83(404), 1023–1032.
- Moon, H. R. and M. Weidner (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83, 1543–1579.
- Pesaran, M. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74(4), 967–1012.
- Pinkse, J., M. E. Slade, and C. Brett (2002). Spatial price competition: a semiparametric approach. Econometrica 70(3), 1111–1153.
- Qu, X. and L. Lee (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. Journal of Econometrics 184(2), 209–232.
- Reich, B. J., M. Fuentes, and D. B. Dunson (2011). Bayesian spatial quantile regression. Journal of the American Statistical Association 106(493), 6–20.
- Shi, W. and L. Lee (2017). Spatial dynamic panel data models with interactive fixed effects. Journal of Econometrics 197, 323–347.
- Stock, J. and M. Watson (2002). Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.
- Villani, M. (2009). Steady-state priors for vector autoregressions. Journal of Applied Econometrics 24(4), 630–650.
- Walker, S. G. and N. L. Hjort (2001). On bayesian consistency. Journal of the Royal Statistical Society: Series B 63, 811–821.
- Yu, J., R. de Jong, and L. Lee (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. Journal of Econometrics 146(1), 118–134.