Bayesian Spatiotemporal Modeling for the Inpatient Hospital Costs of Alcohol-related Disorders

Zhen Yu
Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, China.

Keming Yu
Mathematical Sciences, Brunel University, Uxbridge, UB8 3PH, London, United Kingdom

Wolfgang Karl Härdle
Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

Xueliang Zhang
Department of Medical Engineering and Technology, Xinjiang Medical University, 830054, Urumqi, China.

Kai Wang
Department of Medical Engineering and Technology, Xinjiang Medical University, 830054, Urumqi, China.

Maozai Tian†
Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, China.
Department of Medical Engineering and Technology, Xinjiang Medical University, 830054, Urumqi, China.

Summary. Understanding how healthcare costs vary across different demographics and health conditions is essential to developing policies for healthcare cost reduction. It may not be optimal to apply the conventional mean regression due to its sensitivity to the high level of skewness and spatiotemporal heterogeneity presented in the cost data. To find an alternative method for spatiotemporal analysis with robustness and high estimation efficiency, we combine information across multiple quantiles and propose a Bayesian spatiotemporal weighted composite quantile regression (ST-WCQR) model. An easy-to-implement Gibbs sampling algorithm is provided based on the asymmetric Laplace mixture representation of the error term. Extensive simulation studies show that ST-WCQR outperforms existing methods for skewed error distributions. We apply ST-WCQR to investigate how patients' characteristics affected the inpatient

†Corresponding author. Email: mztian@ruc.edu.cn

hospital costs for alcohol-related disorders and identify areas that could be targeted for cost reduction in New York State from 2015 to 2017.

Keywords Asymmetric Laplace distribution, Bayesian inference, Composite quantile regression, Healthcare cost data, Spatiotemporal model

1. Introduction

According to the U.S. Centers for Medicare & Medicaid (2020), the total health expenditure in the United States grew from $74.1 billion in 1970 to $3.8 trillion in 2019, taking up a rising share of the economy. This is not a surprising phenomenon and is known as Baumol's Cost Disease (Baumol and Bowen, 1965). However, this massive spending could be substantially reduced. Approximately two-thirds of the privately insured patients who showed up in the emergency department (ED) could be treated in the physician offices where the average costs would decrease by 91.8% (UnitedHealth Group, 2019). Moreover, hospitals can improve efficiency in resource allocation after acquiring a good understanding of patterns of hospital costs. Healthcare cost reduction has become more important than ever as the Covid-19 pandemic financial impact is estimated to be $50.7 billion in losses per month for hospitals (American Hospital Association, 2020), and it will continue to be an essential focus in post-pandemic financial recovery. To help develop effective interventions for cost reduction, we examine the spatiotemporal dynamics and the critical determinants of inpatient hospital costs for one of the common reasons for the avoidable ED visits – alcohol-related disorders (nearly 10% of ED visits are avoidable) (Myran et al., 2019).

Geographic variation in healthcare has been documented in a mounting number of studies (Wennberg and Gittelsohn, 1973; Newhouse et al., 2013). Evidence shows that the progressive liberalization of alcohol sales in the United States has led to the unequal availability and heterogeneous patterns of related problems across populations and regions (Connor et al., 2016; Miller et al., 2017; Witkiewitz et al., 2019; Rowell-Cunsolo et al., 2020). From a healthcare policy-making perspective, it is of interest to summarize how the key factors determine the variation in the healthcare outcomes globally (Shoff et al., 2014) since the improvement of population health demands a shared commitment and partnerships across regions (Institute of Medicine (US) Committee on the US Commitment to Global Health, 2009). Such an objective is usually achieved by using global models which assume constant coefficients across space and time to explore the overall covariate effects at a global level (Law et al., 2014; Reilly et al., 2019). As there remains a great amount of heterogeneity after accounting for the disparities in covariates, various types of random intercepts have been considered in the spatiotemporal models with spatiotemporally invariant slopes (Knorr-Held, 2000; Neelon et al., 2015). However, the assumption of homogeneous covariate-response associations cannot provide sufficient information for effective and efficient local interventions (Shoff et al., 2014). It may also overlook some locally spatiotemporal processes, leading to model misspecification for very heterogeneous data (Fotheringham, 1997). To capture the localized covariate-response associations, varying covariate effects may be assumed as a form of heterogeneity (Huang, 2017; Khalili and Chen, 2007) and efforts have been made to develop varying coefficient regression models for the spatiotemporal analysis (Lu et al., 2009; Lee et al., 2021). In the existing spatiotemporal studies, few studies have incorporated both heterogeneity and homogeneity into one framework. For analysis of the areal-referenced alcohol-related hospital costs, we develop a spatiotemporal mixed-effects model with random slopes and random intercepts to help achieve global and local objectives for health improvements and consider variable selection for priority-setting of the healthcare plans due to the limited society's resources.

Given that the healthcare cost data are often characterized by a high level of skewness and spatiotemporal heterogeneity (Gittelsohn and Powe, 1995; Newhouse et al., 2013; Newhouse and Garber, 2013; Neelon et al., 2015; Yang et al., 2019), it remains a challenge to conduct robust and efficient statistical inference for the conditional mean of the costs given possible determinants. Using the ordinary least-square (LS) (Legendre, 1805) estimator or the quadratic loss may not be an optimal choice due to their sensitivity to non-normal errors, outliers, and extreme values (Koenker and Bassett, 1978). Since the pioneering work of Koenker (1984), composite quantile regressions (CQRs) have emerged as a robust and efficient alternative to the mean regression (e.g., via the LS method) in a variety of models. Note that conditional quantile functions are a set of parallel hyperplanes for linear models with error terms independent of the covariates (Koenker and Bassett, 1978). CQR was first proposed to combine multiple quantile regressions (QR) (Koenker and Bassett, 1978) with the equality-of-slopes condition for efficient and robust coefficient estimation in the classical linear model (Koenker, 1984). From a more general modeling perspective, CQR aims to find a set of parallel regression curves and can be viewed as a compromise between a set of quantile regression curves with different slopes and intercepts and a single summary regression curve (Ma and Yin, 2011). Zou and Yuan (2008) found that compared with the LS estimator, CQR has a relative efficiency greater than 70% regardless of the error distribution and sometimes can be arbitrarily more efficient than the LS and single quantile-based methods. For sparse linear models, Huang and Chen (2015) proposed a Bayesian CQR with lasso penalty, which is further developed by Alhamzawi (2016) and Zhao et al. (2016) using other sparsity inducing priors for better performance, especially under the case of heterogeneous errors. Tian et al. (2017) propose a pseudo composite asymmetric Laplace distribution to conduct Bayesian CQR for longitudinal analysis. For semiparametric varying-coefficient partially linear models, Kai et al. (2011) studied the substantial efficiency gain of the proposed CQR estimators for non-normal errors over the LS and single quantile-based techniques. They emphasized that even though the estimation consistency of the CQR estimator breaks down for asymmetric error distributions, the bias term converges to the mean of error, which is assumed to be zero when the number of combined quantile levels is large. Sun et al. (2013) combined quantile regressions with different weights to eliminate the bias terms caused by the asymmetry and propose a weighted composite quantile regression (WCQR) for the local linear regression. They demonstrated the improvements in estimation efficiency of WCQR over equally-weighted CQR and local linear LS estimators for homoscedastic models with asymmetric errors and heteroscedastic models. The idea of CQR could also be found in other statistical problems such as inference for the single-index model (Jiang et al., 2016), regressions with missing data (Luo et al., 2019), and the conditional correlation learning (Ma and Zhang, 2016; Xu, 2017). Though lots of work has been done for CQR, few have investigated its performance in spatiotemporal analysis which has received increasing attention in recent years in fields like health and medical science (Norton and Niu, 2009; Jhuang et al., 2020), environmental science (Reich, 2012; Brynjarsdóttir and Berliner, 2014; Knoblauch and Damoulas, 2018), and criminology (Law et al., 2014; Hu et al., 2018). Given that fitting complicated spatiotemporal structured random effects is challenging via the frequentist methods (Law et al., 2014; Neelon et al., 2015), we study CQR from a Bayesian perspective for the spatiotemporal analysis.

The main contributions of this paper are as follows. First, we develop a spatiotemporal mixed-effects model with random slopes and random intercepts to understand the spatiotemporal patterns of the alcohol-related inpatient hospital costs across seven health service areas of the New York State from 2015 to 2017 and identify the important determinants at both state and local levels. The model not only improves regional estimation from a technical point of view, but also reveals healthcare disparities in different areas, serving as a quantitative reference for the formulation of both state-wide and place-based policies. Second, we develop a continuous shrinkage prior by combining the horseshoe prior and the spatial structure to select the random slopes. It enables us to identify key factors with spatially varying effects while accounting for spatial correlations arising from the neighboring regions. The prior induces a low false positive rate as demonstrated in the simulations, contributing to the priority setting of locally tailored healthcare policies for effective and efficient cost reduction. Third, we propose a Bayesian spatiotemporal weighted composite quantile regression (ST-WCQR) as an alternative to the conventional mean regression. By pooling information across multiple quantiles, ST-WCQR inherits robustness from QR and improves the estimation efficiency compared with the conventional mean regression and single-quantile methods for asymmetric distributions. We verify this through extensive simulation studies via an easy-to-implement Gibbs sampling algorithm. To our knowledge, this is the first study on WCQR in spatiotemporal analysis.

2. Alcohol-related inpatient hospital cost data

As a leading cause of preventable death, alcohol-related disorders, including alcohol intoxication, alcohol use disorder, and alcohol withdrawal, annually cost the United States more than $249 billion (Sacks et al., 2015) and cause approximately 88,000 deaths (Witkiewitz et al., 2019). This paper aims to unveil the spatiotemporal dynamics of alcohol-attributable inpatient hospital costs and their determinants to help formulate interventions for a more efficient resource allocation, related costs reduction, and public health. We concentrate on the alcohol-related inpatients admitted from ED that provides 24-hour costly lifesaving care and serves as a major portal for inpatient admissions (Schuur and Venkatesh, 2012). We obtain the dataset collected by the Statewide Planning and Research Cooperative System (SPARCS) from https://health.data.ny.gov

Table 1. Descriptions of all the variables in the cost data.

Variables	Description
Costs	Inpatients hospital costs ($) for alcohol-related disorders per admission.
Year	The year of discharge: 2015, 2016, 2017.
Areas	Health service areas where hospitals are located: WNY, Finger Lakes, Southern Tier, CNY, Capital/Adirondack, Hudson Valley, NYC.
LOS	Length of stay (days) in a hospital.
Gender	Patient gender: Male (M), Female (F).
Age	Age groups in years at the time of hospital discharge: 0 to 17, 18 to 29, 30 to 49, 50 to 69, 70 or Older.
Race	Patient race: White, Black/African American, Multi-racial, Others.
Ethnicity	Patient ethnicity: Hispanic, Non-Hispanic, Multi-ethnic.
SOI	Severity of illness: Minor, Moderate, Major, Extreme.
ROM	Risk of mortality: Minor, Moderate, Major, Extreme.
Surgical	The APR-DRG classification of medical or surgical: Medical, Surgical.
Payment	Type of payment: Federal insurance programs, Private insurance, Self-pay, Others.

(New York State Department of Health, 2019). It provides de-identified patient-level records of hospital inpatient discharges from 2015 to 2017 for seven health service areas in New York State, i.e., Western New York (WNY), Finger Lakes, Southern Tier, Central New York (CNY), Capital/Adirondack, Hudson Valley, and New York City (NYC).

We preprocess the data set by removing records that have empty entries and combining the detailed payment categories into more general groups, i.e., federal financial health insurance programs (e.g., Medicaid and Medicare), private insurance (e.g., Blue Cross/Blue Shield), self-pay, and others (e.g., department of corrections). The final data set contains 26,448 alcohol-related records across the seven areas over three years.

Table 1 summarizes all the variables used in the analysis. Among these variables, the severity of illness (SOI, the degree of physiologic decompensation or organ system derangement) and the risk of mortality (ROM, the likelihood of dying) are measures assessed through a uniform set of diagnosis-based methods in the All Patient Refined Diagnosis Related Groups payment system used by many US hospitals for inpatient visit classification (Averill et al., 2003). Each patient is assigned to one of the four SOI levels and one of four ROM levels.

Exploratory data analysis shows the characteristics of the inpatients with alcohol-related disorders. The percentage of hospital discharges with an alcohol-related primary admitting diagnosis among all discharges in New York State has steadily grown from 1.99% in 2015 to 2.23% in 2016 and 2.30% in 2017. Among all the discharges with an alcohol-related prin-

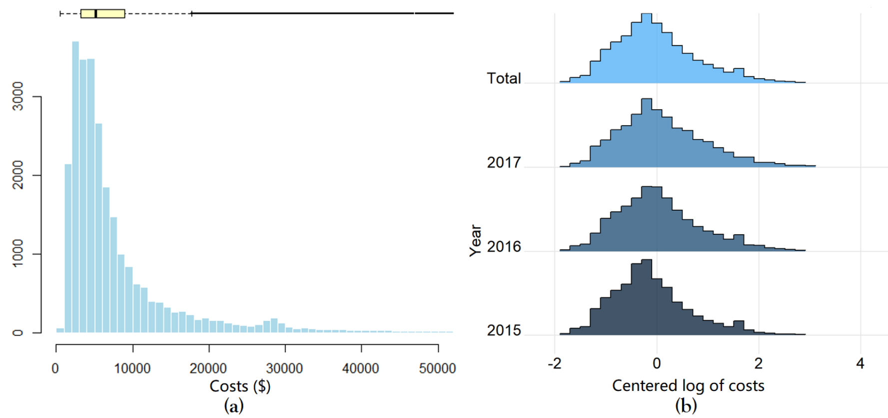

Figure 1. (a) Histogram of the inpatient hospital costs attributable to the alcohol-related disorders. (b) Histograms of the centered log of the alcohol-related costs by year.

cipal admitting diagnosis over the three years, male inpatients accounted for 73.3%. The number of patients was consistently higher for males than females for all age groups (Supplementary Figure D.1). Approximately 88.4% of all the patients admitted with alcohol-related disorders were between the age of 30-69. We also observe that the percentage of alcohol-related hospital admissions with high (major or extreme) ROM and SOI levels increased with age. Moreover, the proportions of patients with high ROM or SOI levels grow more rapidly for males than females as they become 70 or older (Supplementary Figure D.2).

Alcohol-related inpatient hospital costs vary spatiotemporally. With 20% of the patients accounting for 57.5% of the total costs in the New York State, the costs are highly right-skewed as shown in Figure 1(a). The skewness remains even after taking the logarithm of costs (Figure 1(b)). In New York State, the average costs per hospital discharge rose from $7,683.51 in 2015 to $9,412.86 in 2016 and to $10,671.99 in 2017. Average costs of the health service area were the highest in the southeast and lowest in the northeast. Among the seven health service areas, Hudson Valley had the highest number of hospital discharges for patients diagnosed with alcohol-related disorders (5,570, 21.06% of all the related discharges in New York State) and the highest average costs ($12,756.87). Though the discharge counts in NYC (4,743, 17.93%) did not rank high, its average costs ($11,886.40) were the second-highest among the seven areas. The average costs were the lowest in Capital/Adirondack ($5,601.76).

For a deeper understanding of the spatiotemporal dynamics and the driving factors of the alcohol-attributable inpatient hospital costs, we aim to answer the following questions by using a mixed-effects model:

(a) What are the spatiotemporal trends for the inpatient hospital costs of alcohol-related disorders in New York State?

(b) Are there some high-cost areas that could be targeted for interventions to reduce hospital costs?

(c) What are the key state-wide determinants of the hospital costs that require across-region collaborations and efforts to improve the healthcare systems, and how do these factors shape the costs?

(d) Are there any important localized factors of costs that demand carefully tailored place-based policies for healthcare equality?

(e) How to pool the information from nearby spatiotemporal units to improve the small-area estimation? This is critical in the analysis as we observe that the number of cases in the spatiotemporal units (ranging from 32 to 2,930) exhibits a high level of variability.

3. Spatiotemporal weighted composite quantile regression model

Suppose that regions of interest (districts, counties, etc.) are indexed by $i = 1, \ldots, n$ and periods of time (hours, years, etc.) under study by $j = 1, \ldots, J$. Let K_{ij} be the number of the observed cases in region i and period j, $N = \sum_{i=1}^{n} \sum_{j=1}^{J} K_{ij}$ be the total sample size, and $\mathcal{D} = \{y_{ijk}, \boldsymbol{x}_{ijk}\}_{i=1,\ldots,n; j=1,\ldots,J; k=1,\ldots,K_{ij}}$ be the observed data, where y_{ijk} is the continuous response of interest for the k-th subject in region i and period j and $\boldsymbol{x}_{ijk} = (x_{ijk,1}, \ldots, x_{ijk,p})^\top \in \mathbb{R}^p$ is the corresponding p-dimensional covariate vector. Consider a discrete-time linear mixed-effects model specified as

$$\begin{aligned} y_{ijk} &= m(\boldsymbol{x}_{ijk}; \boldsymbol{\beta}, \boldsymbol{\theta}_i, \phi_i, \psi_j, \gamma_{ij}) + \epsilon_{ijk}, \quad \forall i, j, k, \\ m(\boldsymbol{x}_{ijk}; \boldsymbol{\beta}, \boldsymbol{\theta}_i, \phi_i, \psi_j, \gamma_{ij}) &= \boldsymbol{x}_{ijk}^\top (\boldsymbol{\beta} + \boldsymbol{\theta}_i) + \phi_i + \psi_j + \gamma_{ij}, \end{aligned} \tag{1}$$

where the response and the covariates are centered for identifiability in the proposed composite method, $m(\boldsymbol{x}_{ijk})$ represents the conditional mean of the response, and ϵ_{ijk}'s are independent and identically distributed error terms that follow an unknown distribution with mean zero and are independent of covariates. The p-dimensional time-invariant and spatially varying covariate effects are decomposed into two parts, i.e., the mean effects $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ at the global level and the local deviations $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ to the overall mean effects. After adjusting for the covariate effects, we follow Knorr-Held (2000) to divide the spatiotemporal variation into spatial effects ϕ_i, temporal effects ψ_j, and the spatiotemporal interactions γ_{ij} which are three independent unobservable random intercepts. Denote the $n \times p$ random slope matrix as $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)^\top$ with each column representing the spatially varying effects of a covariate. Also, we adopt the following notations: $\boldsymbol{y} = (y_{111}, \ldots, y_{nJK_{nJ}})^\top$, $\boldsymbol{X} = (\boldsymbol{x}_{111}, \ldots, \boldsymbol{x}_{nJK_{nJ}})^\top$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^\top$, $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_J)^\top$, and the vectorized $n \times J$ matrix $\boldsymbol{\gamma}$ composed of γ_{ij} is denoted as $\tilde{\boldsymbol{\gamma}} = \mathrm{vec}(\boldsymbol{\gamma}) = (\gamma_{11}, \ldots, \gamma_{nJ})^\top$. The proposed model (1) includes the model with homogenous covariate effects, $y_{ijk} = \boldsymbol{x}_{ijk}^\top \boldsymbol{\beta} + \phi_i + \psi_j + \gamma_{ij} + \epsilon_{ijk}$, as a special case when $\boldsymbol{\theta}_i = \mathbf{0}, \forall i$. Though temporal heterogeneity in slope could also be considered in the model, we do not include it for the analysis of the alcohol-related cost data as we expect the temporal patterns could be captured through the random intercepts and the spatial pattern of the covariate effects may not change over a short or moderate period of time.

Next, we introduce the CQR to estimate the unknown parameters in model (1). Consider a finite number of quantile levels $0 < \tau_1 < \cdots < \tau_L < 1$. Given that the independence between the error terms and covariates guarantees parallel regression curves (Koenker and Bassett, 1982), the objective function of the proposed ST-WCQR is given by

$$\underset{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Theta}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\gamma}}{\mathrm{Argmin}} \sum_{l=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{K_{ij}} w_l \rho_{\tau_l} \left\{ y_{ijk} - \alpha_l - m(\boldsymbol{x}_{ijk}; \boldsymbol{\beta}, \boldsymbol{\theta}_i, \phi_i, \psi_j, \gamma_{ij}) \right\}, \quad (2)$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L)^\top$ is a vector of quantiles of the error term with respect to $\tau_1, \ldots, \tau_L$, $w_l > 0$ is a quantile-specific weight, $\rho_{\tau_l}(u) = u\{\tau_l - 1(u < 0)\}$ is the quantile-specific check function for $l = 1, \ldots, L$, and $1(\cdot)$ is an indicator function. It is worth pointing out that the conditional mean of the response is what the weighted composite quantile regression aims to estimate by employing a weighted average of check functions with the same coefficient vector across quantiles (Kai et al., 2010; Sun et al., 2013). Using the same coefficients enables the proposed estimator to combine information across different quantiles for estimation efficiency and inherit the robustness from the QR; otherwise, the specification of quantile-dependent coefficients would lead to QRs (Huang and Zhan, 2021). To allow for different amounts of contribution from quantile regression curves to coefficient estimation, quantile-specific weights are employed in (2). When these weights are equal, ST-WCQR reduces to the spatiotemporal CQR that generally leads to less efficiency and less robustness (Bradic et al., 2011; Sun et al., 2013; Jiang et al., 2014; Huang and Chen, 2015; Tian et al., 2017). In addition, ST-WCQR also serves as a unified approach for both mean regression and quantile regression since the ST-WCQR with $L = 1$ reduces to the quantile regression.

Due to the undifferentiability of the check function at point zero, there is no explicit solution to the optimization problem (2). This paper provides a solution from a Bayesian perspective. Suppose that y_{ijk} follows a pseudo composite asymmetric Laplace distribution PCALD$(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\tau})$ (Tian et al., 2017) with probability density function given by $f(y|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\tau}) \propto \prod_{l=1}^{L} \frac{1}{\sigma_l} \exp\{-\rho_{\tau_l}(\frac{y-\mu_l}{\sigma_l})\}$, where $\mu_l = \alpha_l + m(\boldsymbol{x})$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_L)^\top$ is a vector of location parameters, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_L)^\top$ is a vector of scale parameters, and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_L)^\top$ is a vector of skewness parameters. Then, minimizing the objective function (2) is equivalent to maximizing the pseudo-likelihood function

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Theta}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\gamma}, \boldsymbol{\sigma}|\mathcal{D}) = \prod_l \prod_i \prod_j \prod_k \frac{1}{\sigma_l} \exp\left\{-\rho_{\tau_l}\left(\frac{y_{ijk} - \alpha_l - m(\boldsymbol{x}_{ijk})}{\sigma_l}\right)\right\}. \quad (3)$$

In this case, $1/\sigma_1, \ldots, 1/\sigma_L$ serve as weights for the composite method. The asymptotic justification of PCALD could be provided in the same way as ALD in the Bayesian QR (Sriram et al., 2013). However, such a complex likelihood function makes the posterior distribution of $\boldsymbol{\beta}$ analytically untractable. One solution is to use the random walk metropolis algorithm, but parameter tuning is required for the optimal accept rate. Note that PCALD is an extension of ALD which has a mixture representation (Kozumi and Kobayashi, 2011) of $\epsilon \overset{d}{=} \xi V + \sqrt{\zeta \sigma V} Z$, where $\epsilon \sim ALD(\mu, \sigma, \tau)$, μ is a location parameter, $\xi = \frac{1-2\tau}{\tau(1-\tau)}$, $\zeta = \frac{2}{\tau(1-\tau)}$, $V|\sigma \sim \text{Exp}(1/\sigma)$, and $Z \sim \text{N}(0, 1)$. Then, the likelihood function (3) can be decomposed into a hierarchical structure of

$$\begin{cases} L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Theta}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\gamma}, \boldsymbol{v}, \boldsymbol{\sigma}|\mathcal{D}) \\ = \prod_l \prod_i \prod_j \prod_k \frac{1}{\sqrt{2\pi\zeta_l \sigma_l v_{ijk,l}}} \exp\left\{-\frac{(y_{ijk} - \alpha_l - m(\boldsymbol{x}_{ijk}) - \xi_l v_{ijk,l})^2}{2\zeta_l \sigma_l v_{ijk,l}}\right\}, \\ v_{ijk,l}|\sigma_l \sim \text{Exp}(1/\sigma_l), \quad \forall i, j, k, l, \end{cases} \quad (4)$$

where $\boldsymbol{v} = (\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_L^\top)^\top, \boldsymbol{v}_l = (v_{111,l}, \ldots, v_{nJK_{nJ},l})^\top$, $\xi_l = \frac{1-2\tau_l}{\tau_l(1-\tau_l)}$, and $\zeta_l = \frac{2}{\tau_l(1-\tau_l)}$ for $l = 1, \ldots, L$. After assigning appropriate priors to unknown parameters as in Section 4, this mixture representation of the PCALD yields an easy-to-implement Gibbs sampling algorithm for the posterior estimation. Its updating procedure only involves sampling from Gaussian distributions, inverse Gaussian (InvGauss) distributions, and inverse gamma (IG) distributions.

4. Bayesian inference

We adopt intrinsic conditionally autoregressive (ICAR) priors (Besag, 1974) for the random effects to model the spatiotemporal dependency. To be specific, the conditional prior of the spatial effect of region i given all other spatial effects $\boldsymbol{\phi}_{(-i)}$ is specified as $\phi_i|\boldsymbol{\phi}_{(-i)}, \sigma_\phi^2 \sim \mathrm{N}\Big(\frac{1}{b_i}\sum_{i^* \in \partial_i} \phi_{i^*}, \frac{\sigma_\phi^2}{b_i}\Big)$, where ∂_i denotes the surrounding regions of region i, the prior mean is the average spatial effect of the b_i surrounding regions, and σ_ϕ^2 is the conditional variance. Then, under the Brook's Lemma (Banerjee et al., 2003), the joint prior distribution of spatial effects $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^\top$ is

$$\pi(\boldsymbol{\phi}|\sigma_\phi^2) \propto \exp\Big\{-\frac{1}{2\sigma_\phi^2}\boldsymbol{\phi}^\top \boldsymbol{P}\boldsymbol{\phi}\Big\}, \tag{5}$$

where $\boldsymbol{P} = \boldsymbol{B} - \boldsymbol{A}$ is the spatial structure matrix, $\boldsymbol{B} = \mathrm{diag}(b_1, \ldots, b_n)$, and $\boldsymbol{A}$ is the adjacency matrix whose (i_1, i_2)-th entry equals 1 if region i_1 and i_2 are neighbors and 0 otherwise. Since $\boldsymbol{P}\mathbf{1} = \mathbf{0}$, $\boldsymbol{P}$ is singular and the "density" (5) is improper. The common practice to restore propriety is to impose the constraint $\sum_{i=1}^n \phi_i = 0$ via "centering-on-the-fly", i.e., recentering the vector of sampled random effects from the working full conditional distributions around its mean after each MCMC iteration (Banerjee et al., 2003; Norton and Niu, 2009; Neelon et al., 2013).

Similarly, as we expect effects for neighboring periods of time and spatiotemporal units to be alike, we also assign an ICAR prior to $\boldsymbol{\psi}$ and a multivariate ICAR (MICAR) prior to $\tilde{\boldsymbol{\gamma}}$. Then, their joint priors are

$$\pi(\boldsymbol{\psi}|\sigma_\psi^2) \propto \exp\Big\{-\frac{1}{2\sigma_\psi^2}\boldsymbol{\psi}^\top \boldsymbol{R}\boldsymbol{\psi}\Big\}, \tag{6}$$

and

$$\pi(\tilde{\gamma}|\sigma_{\tilde{\gamma}}^2) \propto \exp\Big\{-\frac{1}{2\sigma_{\tilde{\gamma}}^2}\tilde{\gamma}^\top (\boldsymbol{R}\otimes\boldsymbol{P})\tilde{\gamma}\Big\}, \tag{7}$$

where $\otimes$ denotes the Kronecker product, σ_ψ^2 and $\sigma_{\tilde{\gamma}}^2$ are the conditional variances, and the $J \times J$ temporal structure matrix $\boldsymbol{R}$ has entries

$$\boldsymbol{R}_{gu} = \begin{cases} 2 & \text{if } 2 \le g = u \le J-1, \\ 1 & \text{if } g = u = 1 \text{ or } J, \\ -1 & \text{if } |g-u| = 1, \\ 0 & \text{others.} \end{cases} \tag{8}$$

The ICAR prior with the temporal structure matrix R imposed on the temporal effects is actually a first-order random-walk prior since the resulting conditional prior of ψ_j is $\psi_j|\psi_{j-1}, \sigma^2_\psi \sim \mathrm{N}(\psi_{j-1}, \sigma^2_\psi)$ for $j = 2, \ldots, J$ (Rue and Held, 2005). The MICAR prior for $\tilde{\boldsymbol{\gamma}}$ in (7) corresponds to Knorr-Held type IV interaction that arises from the product of the above spatial and temporal structured effects (see Knorr-Held (2000) for further details). Note that there are few observations available for some spatiotemporal units in the alcohol-related cost data. By "borrowing strength" across space and time, these ICAR (MICAR) priors not only help obtain reliable spatiotemporal predictions consistent with nearby regions and periods but also benefit small-area estimation for the analysis. For this reason, these priors are popularly adopted in the existing spatial and spatiotemporal studies (Norton and Niu, 2009; Farnsworth and Ward, 2009; Law et al., 2014; Neelon et al., 2015). Constraints $\sum_{j=1}^{J} \psi_j = 0$ and $\sum_{i=1}^{n} \gamma_{ij} = \sum_{j=1}^{J} \gamma_{ij} = 0$ are imposed via "centering-on-the-fly" in MCMC to make the priors proper (Goicoa, 2018).

We consider variable selection to identify critical factors to the costs, contributing to priority setting in the health care services. The horseshoe prior (Carvalho et al., 2010) is an increasingly commonly used continuous shrinkage prior that outperforms other common scale-mixture priors like Bayesian Lasso (Park and Casella, 2008) and the normal-exponential-gamma prior (Griffin and Brown, 2005). In this paper, we adopt it for the selection of the average effects $\beta_1, \ldots, \beta_p$, which leads to

$$\beta_h|\tau^2_{\boldsymbol{\beta}}, \lambda^2_{\beta_h} \sim N(0, \tau^2_{\boldsymbol{\beta}}\lambda^2_{\beta_h}), \tau^2_{\boldsymbol{\beta}} \sim C^+(0,1), \lambda^2_{\beta_h} \sim C^+(0,1), h = 1, \ldots, p, \quad (9)$$

where $\tau^2_{\boldsymbol{\beta}}$ determines the global shrinkage for all the common effects, $\lambda^2_{\beta_h}$ controls the local shrinkage for β_h, $\Lambda^2_{\boldsymbol{\beta}} = \mathrm{diag}(\lambda^2_{\beta_1}, \ldots, \lambda^2_{\beta_p})$, and $C^+(0,1)$ denotes the standard half-cauchy distribution. Based on the similarity of the shrinkage weights $1 - \kappa_h = 1/(1 + \tau^2_{\boldsymbol{\beta}}\lambda^2_{\beta_h})$ with the posterior inclusion probabilities under the two-groups model, Carvalho et al. (2010) put forward a thresholding rule, i.e., a parameter is considered as a signal if $1 - \hat{\kappa}_h > 0.5$. It attains the Bayes oracle under a 0-1 additive loss up to a multiplicative constant (Datta and Ghosh, 2013). Though other criteria for variable selection (such as credible intervals (van der Pas et al., 2017)) can be used, we find in simulations that the thresholding rule yields a strong control of the false-positive rate to select the non-zero coefficients as demonstrated in Carvalho et al. (2010) and thus adopt it in the costs analysis to avoid the undesirable cases of the insufficient allocation of the limited sources. For conditional conjugacy, we adopt the hierarchical representation of $C^+(0,1)$ (Neville, 2013), which is

$$\begin{aligned} \tau_{\boldsymbol{\beta}}^2|\eta_{\beta_0} &\sim IG(1/2, 1/\eta_{\beta_0}), \quad \eta_{\beta_0} \sim IG(1/2, 1), \\ \lambda_{\beta_h}^2|\eta_{\beta_h} &\sim IG(1/2, 1/\eta_{\beta_h}), \quad \eta_{\beta_h} \sim IG(1/2, 1), \; h = 1, \ldots, p. \end{aligned} \tag{10}$$

Inspired by Mu et al. (2021), we combine the horseshoe prior and the spatial structure matrix to identify which covariates have the spatially correlated effects deviating significantly from the overall mean. Recall that $\boldsymbol{\Theta}_{\cdot h} \in \mathcal{R}^n$, i.e., the h-th column of the random matrix $\boldsymbol{\Theta}$, represents the spatially varying effects of the h-th covariate on the response. We propose a spatial horseshoe prior (SHP) for $\boldsymbol{\Theta}_{\cdot h}$ for $h = 1, \ldots, p$:

$$\pi(\boldsymbol{\Theta}_{\cdot h}|\tau_{\boldsymbol{\Theta}}^2, \lambda_{\boldsymbol{\Theta}_h}^2) \propto \exp\Big\{-\frac{\boldsymbol{\Theta}_{\cdot h}^\top \boldsymbol{P} \boldsymbol{\Theta}_{\cdot h}}{2\tau_{\boldsymbol{\Theta}}^2 \lambda_{\boldsymbol{\Theta}_h}^2}\Big\}, \tau_{\boldsymbol{\Theta}}^2 \sim C^+(0,1), \lambda_{\boldsymbol{\Theta}_h}^2 \sim C^+(0,1), \tag{11}$$

where $\tau_{\boldsymbol{\Theta}}^2$ controls the global shrinkage of all the random slopes, $\lambda_{\boldsymbol{\Theta}_h}^2$ adjusts the local shrinkage for $\boldsymbol{\Theta}_{\cdot h}$, $\Lambda_{\boldsymbol{\Theta}}^2 = \mathrm{diag}(\lambda_{\boldsymbol{\Theta}_1}^2, \ldots, \lambda_{\boldsymbol{\Theta}_p}^2)$, and $\boldsymbol{P}$ is the spatial structure matrix defined in (5). Constraints $\sum_{i=1}^n \Theta_{ih} = 0$ for $h = 1, \ldots, p$ are imposed via "centering-on-the-fly" in MCMC for prior propriety. For variable selection, we use the thresholding rule $1/(1+\hat{\tau}_{\boldsymbol{\Theta}}^2 \hat{\lambda}_{\boldsymbol{\Theta}_h}^2) < 0.5$ like that for the horseshoe prior and demonstrate its satisfactory performance in Section 5. Using the hierarchical representation as in (10), we yield

$$\begin{aligned} \tau_{\boldsymbol{\Theta}}^2|\eta_{\boldsymbol{\Theta}_0} &\sim IG(1/2, 1/\eta_{\boldsymbol{\Theta}_0}), \quad \eta_{\boldsymbol{\Theta}_0} \sim IG(1/2, 1), \\ \lambda_{\boldsymbol{\Theta}_h}^2|\eta_{\boldsymbol{\Theta}_h} &\sim IG(1/2, 1/\eta_{\boldsymbol{\Theta}_h}), \quad \eta_{\boldsymbol{\Theta}_h} \sim IG(1/2, 1), \; h = 1, \ldots, p. \end{aligned} \tag{12}$$

Denote $\boldsymbol{\Omega} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Theta}, \boldsymbol{\phi}, \boldsymbol{\psi}, \tilde{\boldsymbol{\gamma}}, \tau_{\boldsymbol{\beta}}^2, \tau_{\boldsymbol{\Theta}}^2, \boldsymbol{\Lambda}_{\boldsymbol{\beta}}^2, \boldsymbol{\Lambda}_{\boldsymbol{\Theta}}^2, \eta_{\beta_0}, \eta_{\boldsymbol{\Theta}_0}, \{\eta_{\beta_h}, \eta_{\boldsymbol{\Theta}_h}\}_{h=1}^p, \sigma_{\boldsymbol{\phi}}^2, \sigma_{\boldsymbol{\psi}}^2, \sigma_{\tilde{\boldsymbol{\gamma}}}^2, \boldsymbol{v}, \boldsymbol{\sigma}\}$. After assigning the aforementioned priors and independent normal priors to α_l for $l = 1, \ldots, L$ as well as independent inverse gamma priors to σ_l's and random effect variances $\sigma_{\boldsymbol{\phi}}^2$, $\sigma_{\boldsymbol{\psi}}^2$, and $\sigma_{\tilde{\boldsymbol{\gamma}}}^2$, we derive the joint posterior distribution of the unknown parameters as

$$\begin{aligned} \pi(\boldsymbol{\Omega}|\mathcal{D}) \propto & L(\boldsymbol{\Omega}|\mathcal{D})\pi(\boldsymbol{\alpha})\pi(\boldsymbol{v}|\boldsymbol{\sigma})\pi(\boldsymbol{\sigma})\pi(\boldsymbol{\beta}|\tau_{\boldsymbol{\beta}}^2, \boldsymbol{\Lambda}_{\boldsymbol{\beta}}^2)\pi(\tau_{\boldsymbol{\beta}}^2|\eta_{\beta_0})\pi(\eta_{\beta_0})\pi(\boldsymbol{\Lambda}_{\boldsymbol{\beta}}^2|\eta_{\beta_h}) \prod_{h=1}^p \{\pi(\eta_{\beta_h})\pi(\eta_{\boldsymbol{\Theta}_h})\} \times \\ & \pi(\boldsymbol{\Theta}|\tau_{\boldsymbol{\Theta}}^2, \boldsymbol{\Lambda}_{\boldsymbol{\Theta}}^2)\pi(\tau_{\boldsymbol{\Theta}}^2|\eta_{\boldsymbol{\Theta}_0})\pi(\boldsymbol{\Lambda}_{\boldsymbol{\Theta}}^2|\eta_{\boldsymbol{\Theta}_h})\pi(\boldsymbol{\phi}|\sigma_{\boldsymbol{\phi}}^2)\pi(\boldsymbol{\psi}|\sigma_{\boldsymbol{\psi}}^2)\pi(\tilde{\boldsymbol{\gamma}}|\sigma_{\tilde{\boldsymbol{\gamma}}}^2)\pi(\sigma_{\boldsymbol{\phi}}^2)\pi(\sigma_{\boldsymbol{\psi}}^2)\pi(\sigma_{\tilde{\boldsymbol{\gamma}}}^2). \end{aligned} \tag{13}$$

Posterior estimates of parameters are obtained via a Gibbs sampling algorithm with unknown parameters iteratively updated using their full conditional posterior distributions until convergence. The details of the algorithm for ST-WCQR are provided in the supplementary materials.

5. Simulation studies

This section investigates the performance of the ST-WCQR when the error distribution differs from the PCALD under various settings, including the cases of homogeneous and heterogeneous covariate effects, dense and sparse effects, and symmetric and asymmetric errors. For each setting, we examine the ST-WCQR with $L = 1, 3, 5, 9$, among which ST-WCQR with $L = 1$ reduces to the quantile regression for model (1) (denoted by STQR). For comparison, we develop for the spatiotemporal model (1) a conventional mean regression method (STMR) by assuming normal error terms (Lindley and Smith, 1972), multivariate normal prior for $\boldsymbol{\beta}$, and ICAR priors for $\boldsymbol{\Theta}_{\cdot h}$'s (see supplementary Section B.2 for its Gibbs sampling algorithm). We also compare the methods with the spatiotemporal quantile regression method (STQR_Neelon) (Neelon et al., 2015) built for a spatiotemporal model with a common slope vector across space. It is worth mentioning that since the error terms are independent of the covariates, all the methods mentioned above offer estimates for the same quantities and are thus comparable. We set the number of regions and time points as in the cost data and conduct all the simulations in R.

Example 1. *(Homogeneous covariate effects) The data are generated from the model with homogeneous covariate effects across all regions:*

$$y_{ijk} = \boldsymbol{x}_{ijk}\boldsymbol{\beta} + \phi_i + \psi_j + \gamma_{ij} + \epsilon_{ijk},\ \ i = 1, \ldots, 7; j = 1, 2, 3; k = 1, \ldots, 500, \quad (14)$$

where the p-dimensional covariate vectors $\boldsymbol{x}_{ijk}$'s are independently sampled from $N(\mathbf{0}, \boldsymbol{\Sigma})$ with $\Sigma_{h_1 h_2} = 0.5^{|h_1 - h_2|}$ for $h_1, h_2 = 1, \ldots, p$ and the random effects $\boldsymbol{\phi}$, $\boldsymbol{\psi}$, and $\tilde{\boldsymbol{\gamma}}$ are generated from ICAR priors with conditional variances $\sigma^2_\phi = \sigma^2_\psi = \sigma^2_{\tilde{\gamma}} = 2$ and structure matrices $\boldsymbol{P}$ and $\boldsymbol{R}$ of the cost data in Section 2. Simulations are conducted for (a) a dense case with $\boldsymbol{\beta} = \mathbf{1}$, (b) a sparse case with $\boldsymbol{\beta} = (1, -2, \ldots, (-1)^{p/2-1}p/2, 0, \ldots, 0)^\top$, and (c) a very sparse case with $\boldsymbol{\beta} = (1, -2, \ldots, (-1)^{p/4-1}p/4, 0, \ldots, 0)^\top$ for $p = 8, 20$, respectively. Under each setting, we consider heterogeneous error terms $\epsilon_{ijk} = \sigma_{ij}\epsilon^*_{ijk}$ with $\sigma_{ij} \sim Ga(2, 2)$ allowing for spatiotemporal heterogeneity. Choices for the distribution of ϵ^*_{ijk} includes a standard normal distribution $N(0, 1)$, a t distribution $t(3)$, a cauchy distribution $Cauchy(0, 1)$, a log-normal distribution $LN(0, 1)$, a chi-square distribuiton $\chi^2(2)$, and a gamma distribution $Ga(2, 2)$.

Example 2. *(Spatially varying covariate effects) By adding random effects to the regression coefficients, we generate heterogeneous data with* *spatially varying covariate effects under the same settings as Example 1:*

$$y_{ijk} = \boldsymbol{x}_{ijk}(\boldsymbol{\beta} + \boldsymbol{\theta}_i) + \phi_i + \psi_j + \gamma_{ij} + \epsilon_{ijk}. \quad (15)$$

Let $p^* = \#\{\beta_h \neq 0, h = 1, \ldots, p\}$ be the number of nonzero fixed effects. The random slope matrix $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)^\top$ is generated as a sparse matrix whose entries in the first $[p^*/2]$ columns are sampled independently from the ICAR distribution (5) with $\sigma^2_\phi = 2$, where $[\cdot]$ denotes the floor function.

Example 3. *(Sensitivity analysis) We further consider the case when the generation of the spatially correlated coefficients are contaminated:*

$$y_{ijk} = \boldsymbol{x}_{ijk}(\boldsymbol{\beta} + \boldsymbol{\theta}_i + \boldsymbol{u}_i) + \phi_i + \psi_j + \gamma_{ij} + \epsilon_{ijk}, \quad (16)$$

where the spatially correlated (spatially structured) $\boldsymbol{\theta}_i \in \mathcal{R}^n$ is generated from the ICAR distribution (5). Let $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)^\top$ be an $n \times p$ matrix of the unstructured additive effects. Specifically, we set the contamination rate as 20%, i.e., 20% entries (selected randomly) in the first p^* columns of $\boldsymbol{U}$ are independently sampled from the uniform distribution $U(-1, 1)$ and other entries are set as 0's.

We set the quantile levels combined in ST-WCQR as $\tau_l = l/(L+1)$ for $l = 1, \ldots, L$ and use the noninformative prior $\mathrm{IG}(0.001, 0.001)$ for $\sigma^2_{\boldsymbol{\phi}}, \sigma^2_{\boldsymbol{\psi}}, \sigma^2_{\boldsymbol{\gamma}}, \sigma_l$, and the prior $\mathrm{N}(\mu_{0,\alpha_l}, 10^3)$ for α_l with $\mu_{0,\alpha_1}, \ldots, \mu_{0,\alpha_L}$ equally spaced between -1 and 1. All the posterior estimations are obtained through 15,000 MCMC iterations with a burn-in period of 7,000 and a thining parameter of 5. Trace plots and Geweke's z-tests (Geweke, 1992) show that all the chains achieve convergence.

Following Kai et al. (2011), we evaluate the spatiotemporal predictions $\hat{\boldsymbol{y}}$ by the average of median absolute prediction error $\mathrm{MAPE} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{median}\{|\hat{y}^{(t)}_{ijk} - y^{(t)}_{ijk}|, \forall i, j, k\}$ over $T = 20$ simulations, where $\hat{y}^{(t)}_{ijk}$ is the prediction in the t-th simulation run. Coefficient estimation are assessed by mean squared errors (MSEs), i.e., $\mathrm{MSE}^{(\mathrm{stat})} = \frac{1}{pT}\sum_{h=1}^{p}\sum_{t=1}^{T}(\hat{\beta}^{(t)}_h - \beta_h)^2$ for non-zero spatially stationary effects and $\mathrm{MSE}^{(\mathrm{vary})} = \frac{1}{npT}\sum_{i,h,t} (\hat{\beta}^{(t)}_h + \hat{\Theta}^{(t)}_{ih} - \beta_h - \Theta_{ih})^2$ for the spatially varying covariate effects, where $\hat{\beta}^{(t)}_h$'s and $\hat{\Theta}^{(t)}_{ih}$'s are estimates of the fixed and random slopes, respectively. MSEs are also computed for the estimated random intercepts. For example, MSE for the spatial effects is calculated by $\mathrm{MSE}_{\boldsymbol{\phi}} = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}(\hat{\phi}^{(t)}_i - \phi^{(t)}_i)^2$, where $\hat{\phi}^{(t)}_i$ is the estimate of $\phi^{(t)}_i$ in the t-th simulation for all i and t. To conduct variable selection, we use the thresholding rule in Section 4 for ST-WCQR and STQR and 95% credible intervals (CIs) for other methods. We consider the precision=TP/(TP+FP), recall=TP/(TP+FN), and a summarizing criterion F_1-score $= 2 \cdot \text{recall} \cdot \text{precision}/(\text{recall} + \text{precision})$ (Van Rijsbergen, 1979) to evaluate the variable selection procedure, where FN denotes the number of the significant covariates not selected by the method, and TP and FP denote the number of the significant and nonsignificant covariates selected by the method, respectively.

Table 2 compares the performance of ST-WCQR, STMR, STQR, and STQR_Neelon for asymmetric errors when $p = 20$ and $\boldsymbol{\beta} = (1, -2, 3, -4, 5, 0, \ldots, 0)^\top$. Results for other settings are relegated to supplementary materials. Similar results are observed for different choices of p and $\boldsymbol{\beta}$.

We find that for symmetric errors, STQR usually has the best prediction performance among the four methods, and ST-WCQR performs nearly as well as STQR. By taking advantage of useful information across quantiles, ST-WCQR performs comparably to and sometimes better than STQR for variable selection. STQR_Neelon fails to select and estimate the varying coefficients as STQR and ST-WCQR do in Examples 2 & 3 and has the largest prediction errors for data with heterogeneous effects.

However, many real-world data may not be symmetrically distributed and may even be highly skewed like the cost data in Section 2. For asymmetric error distributions, ST-WCQR provides satisfactory performance and has consistently sharp advantages over other methods in terms of spatiotemporal prediction, coefficient estimation, and variable selection, regardless of whether covariate effects are homogeneous or heterogeneous. This suggests that ST-WCQR is robust to heterogeneous effects and non-normal errors. The superiority of the composite method over the conventional mean regression and the single quantile-based counterparts for the asymmetric errors and heterogeneous data is consistent with results in the previous studies for the classical linear regression (Huang and Chen, 2015; Zhao et al., 2016; Alhamzawi, 2016) and longitudinal analysis (Tian et al., 2017, 2021). It is not surprising to find that ST-WCQR has higher estimation efficiency than the single quantile-based STQR and STMR, especially for the random intercepts. This suggests the advantage of pooling information from multiple quantile curves. Moreover, it is worth mentioning that compared with other methods, ST-WCQR has the best performance in selecting both homogeneous covariate effects and spatially varying covariate effects. It has a sharp advantage of controlling the false-positive rate, a finding consistent with Carvalho et al. (2010) for the horseshoe prior in linear regression.

Table 2: The simulation results for parameter estimations, prediction errors, and variable selection by ST-WCQR, STMR, STQR, and STQR_Neelon for the simulated data sets generated from asymmetric error distributions over 20 simulations. Optimal results under each setting are marked in bold.

ϵ	Example	Method	MSE$^{(\text{vary})}$	MSE$^{(\text{stat})}$	MSE			MAPE	$\boldsymbol{\beta}$			$\boldsymbol{\gamma}$		
					$\boldsymbol{\phi}$	$\boldsymbol{\psi}$	$\boldsymbol{\theta}$		precision	recall	F1	precision	recall	F1
LN	1	STQR_Neelon	-	<0.001	0.131	0.041	0.261	0.475	0.977	1.000	0.987	-	-	-
		STQR	-	<0.001	0.132	0.041	0.260	0.473	1.000	0.830	0.906	-	-	-
		STMR	-	0.001	0.366	0.108	0.712	0.747	0.924	1.000	0.957	-	-	-
		ST-WCQR $L = 3$	-	<0.001	0.087	0.027	0.171	**0.441**	1.000	0.860	0.922	-	-	-
		ST-WCQR $L = 5$	-	<0.001	0.072	0.022	0.142	0.442	1.000	0.880	0.933	-	-	-
		ST-WCQR $L = 9$	-	**<0.001**	**0.060**	**0.018**	**0.120**	0.442	1.000	0.980	**0.989**	-	-	-
	2	STQR_Neelon	0.915	<0.001	0.170	0.063	0.331	0.920	0.880	1.000	0.930	-	-	-
		STQR	<0.001	<0.001	0.136	0.057	0.297	0.458	1.000	0.830	0.906	1.000	1.000	**1.000**
		STMR	0.005	0.002	0.364	0.148	0.777	0.722	0.838	1.000	0.905	0.858	1.000	0.910
		ST-WCQR $L = 3$	<0.001	<0.001	0.091	0.038	0.199	0.428	1.000	0.870	0.928	1.000	1.000	**1.000**
		ST-WCQR $L = 5$	<0.001	<0.001	0.075	0.031	0.164	**0.427**	1.000	0.920	0.956	1.000	1.000	**1.000**
		ST-WCQR $L = 9$	**<0.001**	**<0.001**	**0.062**	**0.026**	**0.136**	0.428	1.000	0.950	**0.972**	1.000	1.000	**1.000**
	3	STQR_Neelon	1.008	<0.001	0.165	0.044	0.295	0.940	0.938	1.000	0.965	-	-	-
		STQR	<0.001	<0.001	0.132	0.041	0.259	0.473	1.000	0.830	0.906	1.000	1.000	**1.000**
		STMR	0.006	0.001	0.336	0.108	0.712	0.748	0.924	1.000	0.957	0.850	1.000	0.907
		ST-WCQR $L = 3$	<0.001	<0.001	0.087	0.027	0.171	**0.442**	1.000	0.860	0.922	1.000	1.000	**1.000**
		ST-WCQR $L = 5$	<0.001	<0.001	0.072	0.022	0.143	0.442	1.000	0.880	0.933	1.000	1.000	**1.000**
		ST-WCQR $L = 9$	**<0.001**	**<0.001**	**0.061**	**0.018**	**0.120**	0.443	1.000	0.980	**0.989**	1.000	1.000	**1.000**
χ^2	1	STQR_Neelon	-	<0.001	0.251	0.078	0.493	0.741	0.983	1.000	0.991	-	-	-
		STQR	-	<0.001	0.262	0.080	0.493	0.736	1.000	0.853	0.918	-	-	-
		STMR	-	0.001	0.526	0.162	1.050	0.940	0.899	1.000	0.943	-	-	-
		ST-WCQR $L = 3$	-	<0.001	0.175	0.056	0.333	**0.701**	1.000	0.916	0.953	-	-	-
		ST-WCQR $L = 5$	-	<0.001	0.145	0.046	0.275	0.704	1.000	0.968	0.982	-	-	-
		ST-WCQR $L = 9$	-	**<0.001**	**0.119**	**0.037**	**0.228**	0.703	1.000	0.979	**0.988**	-	-	-
	2	STQR_Neelon	0.911	<0.001	0.303	0.117	0.580	1.124	0.890	1.000	0.944	-	-	-
		STQR	0.002	<0.001	0.260	0.109	0.564	0.713	1.000	0.850	0.917	1.000	1.000	**1.000**
		STMR	0.005	<0.001	0.516	0.220	1.154	0.906	0.900	1.000	0.944	0.825	1.000	0.890
		ST-WCQR $L = 3$	0.001	<0.001	0.177	0.076	0.383	**0.678**	1.000	0.940	0.967	1.000	1.000	**1.000**
		ST-WCQR $L = 5$	0.001	<0.001	0.145	0.063	0.316	0.679	1.000	0.940	0.967	1.000	1.000	**1.000**
		ST-WCQR $L = 9$	**0.001**	**<0.001**	**0.117**	**0.052**	**0.261**	0.679	1.000	0.990	**0.994**	1.000	1.000	**1.000**
	3	STQR_Neelon	1.004	<0.001	0.279	0.084	0.514	1.160	0.894	1.000	0.940	-	-	-
		STQR	0.002	<0.001	0.251	0.078	0.492	0.738	1.000	0.850	0.917	1.000	1.000	**1.000**
		STMR	0.005	0.001	0.526	0.162	1.050	0.940	0.910	0.907	1.000	0.948	0.892	
		ST-WCQR $L = 3$	0.002	<0.001	0.168	0.055	0.332	**0.702**	1.000	0.940	0.967	1.000	1.000	**1.000**
		ST-WCQR $L = 5$	0.001	**<0.001**	0.139	0.044	0.275	0.705	1.000	0.970	0.983	1.000	1.000	**1.000**
		ST-WCQR $L = 9$	**0.001**	<0.001	**0.114**	**0.036**	**0.228**	0.704	1.000	0.980	**0.989**	1.000	1.000	**1.000**
Ga	1	STQR_Neelon	-	**<0.001**	0.092	0.029	0.181	0.310	0.936	1.000	0.964	-	-	-
		STQR	-	<0.001	0.092	0.028	0.181	0.308	1.000	0.810	0.894	-	-	-
		STMR	-	<0.001	0.130	0.040	0.257	0.345	0.855	1.000	0.918	-	-	-
		ST-WCQR $L = 3$	-	<0.001	0.077	0.024	0.151	**0.301**	1.000	0.910	0.950	-	-	-
		ST-WCQR $L = 5$	-	<0.001	0.071	0.022	0.139	0.302	1.000	0.940	0.967	-	-	-
		ST-WCQR $L = 9$	-	<0.001	**0.065**	**0.020**	**0.128**	0.303	1.000	0.980	**0.989**	-	-	-
	2	STQR_Neelon	0.928	<0.001	0.107	0.032	0.202	0.731	0.873	1.000	0.928	-	-	-
		STQR	<0.001	**<0.001**	0.092	0.028	0.181	0.307	1.000	0.810	0.894	1.000	1.000	**1.000**
		STMR	<0.001	<0.001	0.130	0.040	0.257	0.346	0.845	1.000	0.911	0.629	1.000	0.756
		ST-WCQR $L = 3$	<0.001	<0.001	0.077	0.024	0.152	**0.301**	1.000	0.910	0.950	1.000	1.000	**1.000**
		ST-WCQR $L = 5$	<0.001	<0.001	0.071	0.022	0.139	0.303	1.000	0.950	0.972	1.000	1.000	**1.000**
		ST-WCQR $L = 9$	**<0.001**	<0.001	**0.065**	**0.020**	**0.128**	0.302	1.000	0.980	**0.989**	1.000	1.000	**1.000**
	3	STQR_Neelon	1.012	<0.001	0.107	0.032	0.203	0.743	0.860	1.000	0.921	-	-	-
		STQR	<0.001	**<0.001**	0.092	0.028	0.181	0.307	1.000	0.810	0.894	1.000	1.000	**1.000**
		STMR	<0.001	<0.001	0.130	0.040	0.257	0.346	0.845	1.000	0.911	0.629	1.000	0.756
		ST-WCQR $L = 3$	<0.001	<0.001	0.077	0.024	0.151	**0.301**	1.000	0.920	0.956	1.000	1.000	**1.000**
		ST-WCQR $L = 5$	<0.001	<0.001	0.071	0.022	0.139	0.303	1.000	0.950	0.972	1.000	1.000	**1.000**
		ST-WCQR $L = 9$	**<0.001**	<0.001	**0.065**	**0.020**	**0.128**	0.303	1.000	0.980	**0.989**	1.000	1.000	**1.000**

Table 3. Comparison of cross-validated MAPEs for the cost data.

Method	STQR_Neelon	STQR	STMR	ST-WCQR			
				$L = 3$	$L = 5$	$L = 7$	$L = 9$
MAPE	0.279	0.262	0.264	0.251	0.248	0.245	0.244

6. Analysis of the alcohol-related inpatient hospital costs

We apply ST-WCQR to model the relationship between the logarithm of costs and the demographic and health factors after centering the response and standardizing the continuous covariates. Based on the characteristics of the majority of patients, we set the reference group as non-Hispanic (90.19%), white (72.72%), male (73.3%) patients between the age of 50-69 (46.09%) with federal insurance (71.35%), who receive medical treatment (96.48%) with an average length of stay (LOS) (4.9 days) for minor ROM (65.71%) and moderate SOI (52.50%), and who have no spatiotemporal effects. Representing eight categorical variables in Table 1 by dummy variables with the reference group results in twenty-one covariates in the model. Then, we fit the proposed Bayesian ST-WCQR with $L = 1$ (STQR), $3, 5, 7, 9$ for the cost data using the same priors as specified in Section 5, and compare the results with STQR_Neelon (Neelon et al., 2015) and STMR using normally distributed errors (Lindley and Smith, 1972). We run their Gibbs sampling algorithms in R for $15,000$ iterations with a burn-in period of 7,500 and keep every second draw from the sampler for posterior estimation. MCMC convergence is observed by trace plots and Geweke's z-tests (see Supplementary Figure D.3).

In the literature on spatiotemporal analysis, the common ways to evaluate the performance of the models include the time-wise holdout methods (i.e., withhold some observations from the last part of the time series as the testing set and train the model on the remaining observations) (Oliveira et al., 2021; Walker et al., 2022) and the "target-oriented" cross validation (CV) strategies (i.e., variants of CV that deal with either spatial dimensional or temporal dimension or both, which includes leave-location-out CV, leave-time-out CV, and leave-location-and-time-out CV) (Meyer et al., 2018; Gao et al., 2019; Arowosegbe et al., 2022). For the doubly nested spatial structure of the cost data (patients nested within hospitals that are nested within health service areas), we adopt a modified leave-location-out CV to evaluate the performance of the four spatiotemporal regression methods. We use the facility ID assigned to each hospital by the New York State Department (which is provided in

Table 4. Comparison of the results for $\hat{\beta}$ and their CIs in parenthesis by ST-WCQR with $L = 9$, STMR, STQR, and STQR_Neelon.

Covariates	STQR_Neelon	STQR	STMR	ST-WCQR
LOS	$0.629_{(0.619,0.639)}$	$0.697_{(0.683,0.709)}$	$0.494_{(0.486,0.503)}$	$0.641_{(0.636,0.646)}$
Extreme ROM	$0.365_{(0.311,0.413)}$	$0.361_{(0.313,0.411)}$	$0.427_{(0.372,0.484)}$	$0.366_{(0.347,0.384)}$
Extreme SOI	$0.339_{(0.295,0.388)}$	$0.281_{(0.236,0.324)}$	$0.393_{(0.343,0.443)}$	$0.313_{(0.297,0.330)}$
Major ROM	$0.258_{(0.230,0.287)}$	$0.257_{(0.229,0.284)}$	$0.308_{(0.274,0.341)}$	$0.268_{(0.257,0.278)}$
Surgical	$0.215_{(0.184,0.247)}$	$0.183_{(0.154,0.212)}$	$0.209_{(0.172,0.246)}$	$0.200_{(0.188,0.212)}$
Multi-ethnic	$0.432_{(0.255,0.608)}$	$0.215_{(-0.117,0.481)}$	$0.181_{(-0.194,0.500)}$	$0.197_{(-0.043,0.385)}$
Major SOI	$0.217_{(0.198,0.235)}$	$0.189_{(0.170,0.206)}$	$0.223_{(0.201,0.244)}$	$0.190_{(0.183,0.197)}$
Moderate ROM	$0.183_{(0.167,0.198)}$	$0.168_{(0.153,0.183)}$	$0.175_{(0.155,0.194)}$	$0.173_{(0.167,0.179)}$
Minor SOI	$-0.108_{(-0.121,-0.095)}$	$-0.115_{(-0.129,-0.101)}$	$-0.159_{(-0.178,-0.141)}$	$-0.141_{(-0.146,-0.135)}$
Age 18-29	$-0.088_{(-0.108,-0.067)}$	$-0.094_{(-0.115,-0.075)}$	$-0.107_{(-0.136,-0.081)}$	$-0.112_{(-0.119,-0.104)}$
Other races	$0.112_{(0.091,0.137)}$	$0.065_{(0.043,0.088)}$	$0.044_{(0.014,0.075)}$	$0.102_{(0.090,0.114)}$
Others payments	$0.088_{(0.041,0.135)}$	$0.083_{(0.035,0.131)}$	$0.116_{(0.060,0.172)}$	$0.095_{(0.078,0.111)}$
Multi-racial	$0.118_{(0.034,0.201)}$	$0.081_{(-0.029,0.205)}$	$0.122_{(-0.052,0.300)}$	$0.067_{(-0.009,0.148)}$
Age 30-49	$-0.053_{(-0.065,-0.042)}$	$-0.056_{(-0.068,-0.045)}$	$-0.059_{(-0.074,-0.045)}$	$-0.062_{(-0.067,-0.058)}$
Self-pay	$0.019_{(-0.001,0.041)}$	$-0.019_{(-0.04,0.001)}$	$-0.028_{(-0.057,0.001)}$	$-0.027_{(-0.036,-0.017)}$
Age 0-17	$0.053_{(-0.079,0.184)}$	$0.038_{(-0.052,0.166)}$	$0.018_{(-0.149,0.178)}$	$0.024_{(-0.018,0.075)}$
Black/African American	$0.026_{(0.008,0.047)}$	$0.008_{(-0.007,0.025)}$	$-0.007_{(-0.031,0.018)}$	$0.021_{(0.013,0.031)}$
Hispanic	$0.072_{(0.047,0.099)}$	$0.017_{(-0.011,0.045)}$	$0.014_{(-0.028,0.055)}$	$-0.015_{(-0.034,0.002)}$
Female	$-0.012_{(-0.024,-0.001)}$	$-0.013_{(-0.025,-0.001)}$	$-0.013_{(-0.028,0.002)}$	$-0.014_{(-0.019,-0.009)}$
Age ≥ 70	$0.016_{(-0.015,0.046)}$	$-0.005_{(-0.029,0.018)}$	$0.005_{(-0.027,0.037)}$	$-0.005_{(-0.016,0.004)}$
Private insurance	$0.014_{(0.001,0.028)}$	$-0.001_{(-0.013,0.012)}$	$0.006_{(-0.012,0.023)}$	$0.003_{(-0.002,0.009)}$

the cost data) to divide the spatiotemporal observations into 10 folds of approximately the same size, ensuring that all the discharge records from one hospital would belong to the same fold. This suggests that the future information of one location would not be used to predict the past in the CV. Then, each fold is iteratively used as the testing set to compute MAPE with models trained on the data from the remaining fold. The average MAPEs over the 10 folds for each model are summarized in Table 3. It is shown that ST-WCQRs have better prediction performance than other methods and ST-WCQR with $L = 9$ performs the best. We also employ a holdout method with 10% of the total observations randomly withheld from the last year for model validation. The results are quite similar to those in Table 3. In addition, we study the sensitivity by changing the hyperparameters of the priors and find the results from ST-WCQR with $L = 9$ are quite robust to the choice of these priors (results are not shown for saving space).

Table 4 compares the estimates of overall average effects $\boldsymbol{\beta}$ and their 95% CIs from the four spatiotemporal methods based on the full dataset. As the horseshoe prior is better at suppressing noise than many other priors and leaving obvious signals unshrunk (Carvalho et al., 2010), it is expected that for median regression methods, STQR usually gives closely matched large coefficient estimates to STQR_Neelon estimates and shrunk results for smaller coefficient estimates. Furthermore, we find that the CIs provided by STQR are often narrower than those by STQR_Neelon. As for regression of the conditional mean of the costs, all the CIs estimated by ST-WCQR with $L = 9$ are much narrower than those of STMR. This suggests that a more precise estimation can be provided by ST-WCQR.

By using ST-WCQR with $L = 9$ and the thresholding rule mentioned in Section 4, LOS and Extreme ROM are identified to have the statistically significant state-wide effects on the average costs. This finding is supported by the previous studies documenting that LOS and acute clinical features are strongly associated with the increased inpatient costs (Fine et al., 2000; Wei et al., 2010). More precisely, an additional day in hospital for the reference patient group increases the average costs on the original scale by 11.16% (95% CI [11.07%, 11.24%]) (see a detailed discussion on the interpretation of the coefficients in the log-linear regression in Halvorsen and Palmquist (1980)). This suggests that efforts to reduce LOS, for example, through discharge planning (Wei et al., 2015) and inpatient addiction consult service (Weinstein et al., 2018), may be positive steps toward effectively lowering the alcohol-related hospital costs. Costs also surge by 44.16% (95% CI [41.54%, 46.84%]) for inpatients at extreme ROM, which highlights the need to promote early entry to treatment. As it is found that many individuals who suffer from alcohol-related disorders are reluctant to seek treatment until they experience emergent or advanced illness (Tuithof et al., 2016; Connor et al., 2016), interventions that promote early entry to treatment before the disorders are well developed may help reduce the population burden of the alcohol-related harms. For example, such promotion can be achieved by brief behavioral intervention (Connor et al., 2016) and systematic screening (Carvalho et al., 2019) in primary care. Though STMR identifies these two covariates as significant factors as well, ST-WCQR yields a sparser model with greater interpretability and produces more precise coefficient estimates with narrower CIs. These findings are of great importance for knowledge-based priority settings in healthcare plans. From a policy-making perspective, the ST-WCQR significantly narrows down the possible focuses of intervention to efficient ones and highlights the importance of the state-wide joint efforts towards improved hospital management for effective hospital length of stay as well as efficient patient triage and resource allocation in the ED for the extreme-risk patients.

Furthermore, there is a statistically significant spatial heterogeneity in hospital costs for Multi-racial patients and Multi-ethnic patients. This finding reinforces previous studies that there exists a difference in health care spending by minority groups (Dieleman et al., 2021) and this dispar-

Table 5. The estimated significant spatially varying effects and their CIs from ST-WCQR with $L = 1$ (STQR) and $L = 9$ by the thresholding rule.

Region	Multi-racial		Multi-ethnic
Method	ST-WCQR	STQR	ST-WCQR
WNY	$-0.867_{(-1.060,-0.668)}$	$-0.784_{(-1.147,-0.464)}$	$0.073_{(-0.408,0.517)}$
Finger Lakes	$-0.103_{(-0.287,0.094)}$	$-0.138_{(-0.363,0.120)}$	$0.16_{(-0.283,0.610)}$
Southern Tier	$-0.026_{(-1.029,0.947)}$	$-0.016_{(-0.981,0.913)}$	$-0.017_{(-0.878,0.794)}$
CNY	$-0.083_{(-0.274,0.113)}$	$-0.077_{(-0.330,0.188)}$	$0.052_{(-1.441,1.363)}$
Capital/Adirondack	$0.974_{(0.787,1.178)}$	$0.969_{(0.735,1.247)}$	$-0.146_{(-1.432,1.173)}$
Hudson Valley	$-0.069_{(-0.287,0.163)}$	$-0.061_{(-0.440,0.247)}$	$-0.273_{(-0.749,0.164)}$
NYC	$0.174_{(-0.271,0.558)}$	$0.106_{(-0.328,0.576)}$	$0.152_{(-0.519,0.820)}$

ity is further quantified by ST-WCQR as summarized in Table 5. To help acquire the knowledge of their geographical patterns, additional figures that map these estimates are also provided in the supplements (Figure D.4). A much greater disparity in average costs for multi-racial patients can be observed across health service areas when compared with costs for Multi-ethnic patients. For example, compared with the reference group, there is a striking increase (by 97.4%, 95% CI (78.7%,117.8%)) in average costs for multi-racial patients in Capital and a much lower average costs (decrease by 86.7%, 95% CI (66.8%,106.0%)) for multi-racial patients in WNY. This spatial heterogeneity corresponds to the key location-specific needs required to address in the locally tailored health policy for improved effectiveness of policy implementation and healthcare equality. More effort is needed to explore the drivers that shape these disparities. For median regression of the costs, STQR (ST-WCQR with $L = 1$) also identifies spatial heterogeneity in the median costs of multi-racial patients, which STQR_Neelon fails to capture because of its spatial homogeneity assumption on the slopes. The heterogeneous nature of the effects revealed by STQR may help explain why STQR_Neelon produces wider CIs for the multiracial coefficient than many other covariate coefficients in Table 4.

We display the median of the true and estimated alcohol-related costs for each area in Figure 2(a). ST-WCQR captures the spatiotemporal characteristics of costs well, especially for 2015 and 2016. There is a slight deviation in the estimation from the true value in 2017 because the total sample size in 2017 is relatively small and is only one-third of the sample size in both 2015 and 2016.

To reveal the disparities in costs across space and time, we also plot the yearly spatiotemporal random effects (STRE) $\phi_i + \psi_j + \gamma_{ij}$, which represent the deviation in the average costs relative to the reference level

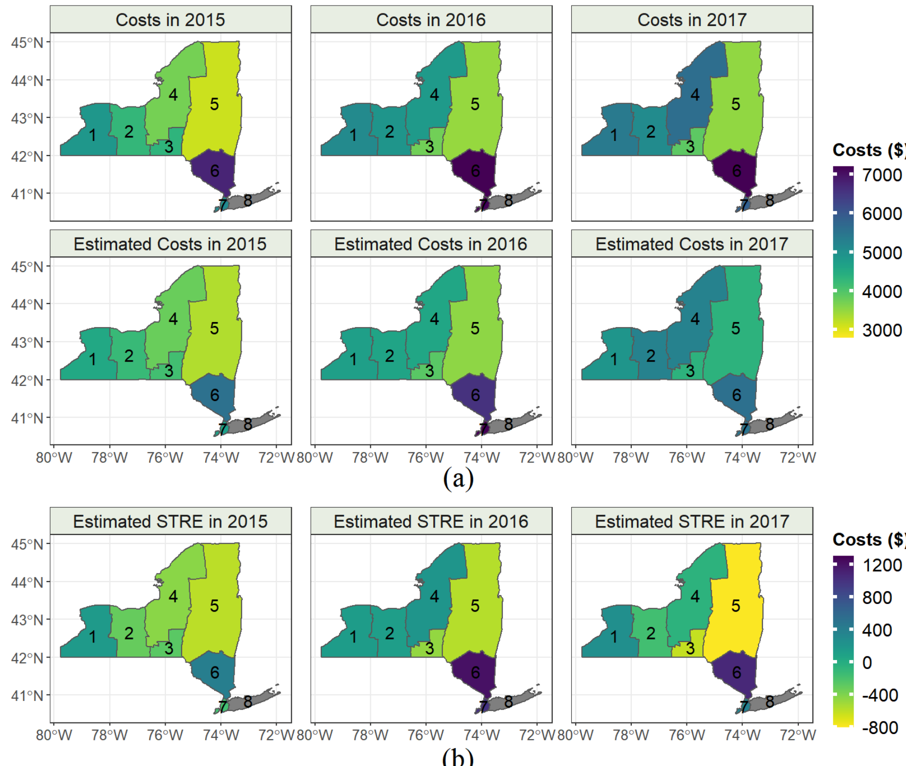

Figure 2. (a) The median of true and estimated (by ST-WCQR with $L = 9$) alcohol-attributable hospital inpatients costs on the original scale in seven health service areas of New York State from 2015 to 2017. (b) The estimated spatiotemporal random effects (STRE), $\hat{\phi}_i + \hat{\psi}_j + \hat{\gamma}_{ij}$, over the three years by ST-WCQR with $L = 9$. Since there is no data for Long Island, we color the area in grey. Areas are labeled as 1: WNY, 2: Finger Lakes, 3: Southern Tier, 4: CNY, 5: Capital/Adirondack, 6: Hudson Valley, and 7: NYC.

($5718.67) across space and time in Figure 2(b). It exhibits some degree of spatial heterogeneity and local temporal trends in the average costs that covariates can not explain, supporting the use of the spatiotemporal model in this application and also suggesting a need for area-specific healthcare costs intervention. More precisely, average costs in the darkest area (Hudson Valley in 2016) and the lightest area (Capital/Adirondack in 2017) of these plots were $1210.57 higher and $792.84 lower than the reference level which was estimated to be $5718.67 per admission on average, respectively. Western areas had relatively stationary STRE values over the three years, while it was not the case in the east. Average costs in Capital/Adirondack remained at round \$570 lower than the average costs for the reference group during 2015 and 2016 and the difference was further widened as the average costs in Capital/Adirondack in 2017 suddenly dropped by \$223.64. On the contrary, the average costs in both Hudson Valley and NYC surged by over \$800 in 2016, exceeding the reference level by approximately \$1,000. Their cost averages remained higher than the reference level at \$1067.22 for Hudson Valley and \$373.65 for New York City, making them ideal targets for local intervention to reduce alcohol-related costs.

7. Discussion

It is challenging to understand the key drivers of hospital costs and the spatiotemporal patterns under the heterogeneity and high level of skewness. In this paper, we propose a robust and efficient alternative to the conventional mean regression that summarizes the spatiotemporal covariate effects on the conditional mean of the alcohol-related inpatient hospital costs. By combining information across quantiles, we propose a Bayesian ST-WCQR model with spatially varying random slopes and spatiotemporal random intercepts and adopt continuous shrinkage priors to select the important covariates. ST-WCQR enables the investigation of both the important region-wide covariate effects and the heterogeneous ones across regions, providing a quantitative reference for the priority setting of the multi-level healthcare policies. Meanwhile, it can also be used to identify areas with fast-changing high hospital costs for cost reduction. Extensive simulation studies show that compared with STMR, STQR, and STQR_Neelon (Neelon et al., 2015), ST-WCQR has comparable prediction performance to the best method for symmetric error distributions and is superior in coefficients estimation and prediction performance with a low level of the false-positive rate for asymmetric error distributions. Moreover, ST-WCQR can be viewed as a unified approach to obtaining the best prediction performance for both symmetric and asymmetric errors since STQR is a special case of ST-WCQR when $L = 1$. The method is also applicable to monitoring the spatiotemporal variation of the healthcare costs during and after the Covid-19 pandemic once the relevant data becomes available.

It is worth emphasizing that this paper aims to develop a robust and efficient estimator in spatiotemporal analysis for inference on the conditional mean of the continuous response. This is accomplished by holding the covariate coefficients across quantiles to be the same. The method could also serve as a starting point for robust spatiotemporal statistical inference beyond the mean regression. Given that there may be shared information among quantile-specific coefficients across neighboring quantile levels, the equality-of-slopes condition could be imposed on several different continuous intervals of quantile levels to accommodate the commonality. This would yield robust and efficient spatiotemporal quantile regression estimators, contributing to a more refined healthcare policy formulation for the high-cost and low-cost populations. Moreover, the proposed method can also be adapted to model the spatiotemporal trends of discrete variables, such as disease case counts and counts of crime.

Acknowledgments

The authors thank the Joint Editor, an Associate Editor, and two anonymous reviewers for their constructive comments and suggestions. Dr. Zhen Yu's research was supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (21XNH155). Prof. Maozai Tian's work was partially supported by the National Natural Science Foundation of China (No.11861042), and the China Statistical Research Project (No.2020LZ25). This research was also supported by Public Computing Cloud, Renmin University of China.

References

Alhamzawi, R. (2016) Bayesian analysis of composite quantile regression. *Stat. Biosci.*, **8**, 358–373.

American Hospital Association (2020) Hospitals and health systems face unprecedented financial pressures due to covid-19. American Hospital Association, Chicago (Available from `https://www.aha.org/system/files/media/file/2020/05/aha-covid19-financial-impact-0520-FINAL.pdf`).

Arowosegbe, O. O., Röösli, M., Künzli, N., Saucy, A., Adebayo-Ojo, T. C., Schwartz, J., Kebalepile, M., Jeebhay, M. F., Dalvie, M. A. and de Hoogh, K. (2022) Ensemble averaging using remote sensing data to model spatiotemporal pm10 concentrations in sparsely monitored south africa. *Environmental Pollution*, **310**, 119883.

Averill, R. F., Goldfield, N., Hughes, J. S., Bonazelli, J., McCullough, E. C., Steinbeck, B. A., Mullin, R., Tang, A. M., Muldoon, J., Turner, L. and Gay, J. (2003) All patient refined diagnosis related groups (APR-DRGs): Methodology overview. Wallingford, CT: 3M Health Information Systems (Available from `https://www.hcup-s.ahrq.gov/db/nation/nis/APR-DRGsV20MethodologyOverviewandBibliography.pdf`).

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2003) *Hierarchical modeling and analysis for spatial data.* Boca Raton: Chapman & Hall/CRC.

Baumol, W. J. and Bowen, W. G. (1965) On the performing arts: The anatomy of their economic problems. *Am. Econ. Rev.*, **55**, 495–502.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc. B*, **36**, 192–225.

Bradic, J., Fan, J. and Wang, W. (2011) Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *J. R. Stat. Soc., B: Stat. Methodol.*, **73**, 325–349.

Brynjarsdóttir, J. and Berliner, L. M. (2014) Dimension-reduced modeling of spatio-temporal processes. *J. Am. Stat. Assoc.*, **109**, 1647–1659.

Carvalho, A. F., Heilig, M., Perez, A., Probst, C. and Rehm, J. (2019) Alcohol use disorders. *Lancet*, **394**, 781–792.

Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010) The horseshoe estimator for sparse signals. *Biometrika*, **97**, 465–480.

Connor, J. P., Haber, P. S. and Hall, W. D. (2016) Alcohol use disorders. *Lancet*, **387**, 988–998.

Datta, J. and Ghosh, J. K. (2013) Asymptotic Properties of Bayes Risk for the Horseshoe Prior. *Bayesian Anal.*, **8**, 111 – 132.

Dieleman, J. L., Chen, C., Crosby, S. W., Liu, A., McCracken, D., Pollock, I. A., Sahu, M., Tsakalos, G., Dwyer-Lindgren, L., Haakenstad, A., Mokdad, A. H., Roth, G. A., Scott, K. W. and Murray, C. J. L. (2021) US Health Care Spending by Race and Ethnicity, 2002-2016. *JAMA*, **326**, 649–659.

Farnsworth, M. L. and Ward, M. P. (2009) Identifying spatio-temporal patterns of transboundary disease spread: examples using avian influenza H5N1 outbreaks. *Vet. Res.*, **40**, 20.

Fine, M. J., Pratt, H. M., Obrosky, D. S., Lave, J. R., McIntosh, L. J., Singer, D. E., Coley, C. M. and Kapoor, W. N. (2000) Relation between length of hospital stay and costs of care for patients with community-acquired pneumonia. *Am. J. Med.*, **109**, 378–385.

Fotheringham, A. S. (1997) Trends in quantitative methods i: stressing the local. *Prog. Hum. Geogr.*, **21**, 88–96.

Gao, J., Liang, T., Yin, J., Ge, J., Feng, Q., Wu, C., Hou, M., Liu, J. and Xie, H. (2019) Estimation of alpine grassland forage nitrogen coupled with hyperspectral characteristics during different growth periods on the tibetan plateau. *Remote Sensing*, **11**.

Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics* (eds. A. P. D. J. M. Bernardo, J. O. Berger and A. F. M. Smith), 169–194. New York: Oxford Press.

Gittelsohn, A. and Powe, N. (1995) Small area variations in health care delivery in maryland. *Health Serv. Res.*, **30**, 295—317.

Goicoa, T., A. A. U. M. D. . H. J. S. (2018) In spatio-temporal disease mapping models, identifiability constraints affect pql and inla results. *Stoch. Environ. Res. Risk Assess.*, **32**, 749–770.

Griffin, J. and Brown, P. (2005) Alternative prior distributions for variable selection with very many more variables than observations. *Tech. rep.*, Technical report, University of Warwick.

Halvorsen, R. and Palmquist, R. (1980) The interpretation of dummy variables in semilogarithmic equations. *Am. Econ. Rev.*, **70**, 474–475.

Hu, T., Zhu, X., Duan, L. and Guo, W. (2018) Urban crime prediction based on spatio-temporal bayesian model. *PLoS One*, **13**, 1–18.

Huang, H. (2017) Regression in heterogeneous problems. *Stat. Sin.*, **27**, 71–88.

Huang, H. and Chen, Z. (2015) Bayesian composite quantile regression. *J. Stat. Comput. Simul.*, **85**, 3744–3754.

Huang, X. and Zhan, Z. (2021) Local composite quantile regression for regression discontinuity. *J. Bus. Econ. Stat.*, **0**, 1–13.

Institute of Medicine (US) Committee on the US Commitment to Global Health (2009) *The US commitment to global health: recommendations for the public and private sectors.* Washington (DC): National Academies Press (US).

Jhuang, A., Fuentes, M., Bandyopadhyay, D. and Reich, B. (2020) Spatiotemporal signal detection using continuous shrinkage priors. *Statist. Med.*, **39**, 1–16.

Jiang, J., Jiang, X. and Song, X. (2014) Weighted composite quantile regression estimation of dtarch models. *Econom. J.*, **17**, 1–23.

Jiang, R., Qian, W.-M. and Zhou, Z.-G. (2016) Weighted composite quantile regression for single-index models. *J. Multivar. Anal.*, **148**, 34–48.

Kai, B., Li, R. and Zou, H. (2010) Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *J. R. Stat. Soc., B: Stat.*, **72**, 49–69.

— (2011) New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Ann. Stat.*, **39**, 305 – 332.

Khalili, A. and Chen, J. (2007) Variable selection in finite mixture of regression models. *Journal of the American Statistical association*, **102**, 1025–1038.

Knoblauch, J. and Damoulas, T. (2018) Spatio-temporal bayesian online changepoint detection with model selection. In *Proceedings of the 35th International Conference on Machine Learning* (eds. J. Dy and A. Krause), vol. 80 of *Proceedings of Machine Learning Research*, 2718–2727. PMLR.

Knorr-Held, L. (2000) Bayesian modelling of inseparable space-time variation in disease risk. *Statist. Med.*, **19**, 2555–2567.

Koenker, R. (1984) A note on l-estimates for linear models. *Stat. Probab. Lett.*, **2**, 323–325.

Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica*, **46**, 33–50.

Koenker, R. and Bassett, J. G. (1982) Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, 43–61.

Kozumi, H. and Kobayashi, G. (2011) Gibbs sampling methods for bayesian quantile regression. *J. Stat. Comput. Simul.*, **81**, 1565–1578.

Law, J., Quick, M. and Chan, P. (2014) Bayesian spatio-temporal modeling for analysing local patterns of crime over time at the small-area level. *J. Quant. Criminol.*, **30**, 57–78.

Lee, J., Kamenetsky, M. E., Gangnon, R. E. and Zhu, J. (2021) Clustered spatio-temporal varying coefficient regression model. *Stat. Med.*, **40**, 465–480.

Legendre, A. M. (1805) *Nouvelles méthodes pour la détermination des orbites des comètes.* Courcier: Paris: F. Didot.

Lindley, D. V. and Smith, A. F. (1972) Bayes estimates for the linear model. *J. R. Stat. Soc., B: Stat. Methodol.*, **34**, 1–18.

Lu, Z., Steinskog, D. J., Tjøstheim, D. and Yao, Q. (2009) Adaptively varying-coefficient spatiotemporal models. *J. R. Stat. Soc., B: Stat. Methodol.*, **71**, 859–880.

Luo, S., Zhang, C.-y. and Wang, M. (2019) Composite quantile regression for varying coefficient models with response data missing at random. *Symmetry*, **11**, 1065.

Ma, X. and Zhang, J. (2016) Robust model-free feature screening via quantile correlation. *J. Multivar. Anal.*, **143**, 472–480.

Ma, Y. and Yin, G. (2011) Censored quantile regression with covariate measurement errors. *Stat. Sin.*, **21**, 949.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. and Nauss, T. (2018) Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, **101**, 1–9. URL: https://www.sciencedirect.com/science/article/pii/S1364815217310976.

Miller, T. R., Nygaard, P., Gaidus, A., Grube, J. W., Ponicki, W. R., Lawrence, B. A. and Gruenewald, P. J. (2017) Heterogeneous costs of alcohol and drug problems across cities and counties in california. *Alcohol. Clin. Exp. Res.*, **41**, 758—768.

Mu, J., Liu, Q., Kuo, L. and Hu, G. (2021) Bayesian variable selection for the cox regression model with spatially varying coefficients with applications to louisiana respiratory cancer data. *Biom. J.*, **63**, 1607–1622.

Myran, D. T., Hsu, A. T., Smith, G. and Tanuseputro, P. (2019) Rates of emergency department visits attributable to alcohol use in ontario from 2003 to 2016: a retrospective population-level study. *Can. Med. Assoc. J.*, **191**, E804–E810.

Neelon, B., Ghosh, P. and Loebs, P. F. (2013) A spatial poisson hurdle model for exploring geographic variation in emergency department visits. *J. R. Statist. Soc. A*, **176**, 389–413.

Neelon, B., Li, F., Burgette, L. F. and Benjamin Neelon, S. E. (2015) A spatiotemporal quantile regression model for emergency department expenditures. *Statist. Med.*, **34**, 2559–2575.

Neville, S. E. (2013) *Elaborate distribution semiparametric regression via mean field variational Bayes*. Ph.D. thesis, University of Wollongong.

New York State Department of Health (2019) Hospital inpatient discharges (SPARCS de-identified): 2015-2017. New York State Department of Health, New York (Available from https://health.data.ny.gov/Health/Hospital-Inpatient-Discharges-SPARCS-De-Identified/82xm-y6g8, https://health.data.ny.gov/Health/Hospital-Inpatient-Discharges-SPARCS-De-Identified/gnzp-ekau, and https://health.data.ny.gov/dataset/Hospital-Inpatient-Discharges-SPARCS-De-Identified/22g3-z7e7).

Newhouse, J. P., Alan, M. G., Robin, P. G., Margaret, A. M., Michelle, M. and Ashna, K. e. (2013) Variation in health care spending: Target decision making, not geography. Washington, DC: The National Academies Press.

Newhouse, J. P. and Garber, A. M. (2013) Geographic Variation in Health Care Spending in the United States: Insights From an Institute of Medicine Report. *JAMA*, **310**, 1227–1228.

Norton, J. D. and Niu, X.-F. (2009) Intrinsically autoregressive spatiotemporal models with application to aggregated birth outcomes. *J. Am. Stat. Assoc.*, **104**, 638–649.

Oliveira, M., Torgo, L. and Santos Costa, V. (2021) Evaluation procedures for forecasting with spatiotemporal data. *Mathematics*, **9**, 691.

Park, T. and Casella, G. (2008) The bayesian lasso. *J. Am. Stat. Assoc.*, **103**, 681–686.

van der Pas, S., Szabó, B. and van der Vaart, A. (2017) Uncertainty quantification for the horseshoe (with discussion). *Bayesian Anal.*, **12**, 1221–1274.

Reich, B. J. (2012) Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *J. R. Statist. Soc. C*, **61**, 535–553.

Reilly, K. H., Bartley, K., Paone, D. and Tuazon, E. (2019) Alcohol-related emergency department visits and income inequality in New York City, USA: an ecological study. *Epidemiology and health*, **41**.

Rowell-Cunsolo, T. L., Liu, J., Hu, G. and Larson, E. (2020) Length of hospitalization and hospital readmissions among patients with substance use disorders in New York City, NY USA. *Drug Alcohol Depend.*, **212**, 107987.

Rue, H. and Held, L. (2005) *Gaussian Markov random fields: theory and applications.* No. 104 in Monographs on statistics and applied probability. Boca Raton: Chapman & Hall/CRC.

Sacks, J. J., Gonzales, K. R., Bouchery, E. E., Tomedi, L. E. and Brewer, R. D. (2015) 2010 national and state costs of excessive alcohol consumption. *Am. J. Prev. Med.*, **49**, e73–e79.

Schuur, J. D. and Venkatesh, A. K. (2012) The growing role of emergency departments in hospital admissions. *N. Engl. J. Med.*, **367 5**, 391–393.

Shoff, C., Chen, V. Y.-J. and Yang, T.-C. (2014) When homogeneity meets heterogeneity: the geographically weighted regression with spatial lag approach to prenatal care utilisation. *Geospat. Health*, **8**, 557–568.

Sriram, K., Ramamoorthi, R. and Ghosh, P. (2013) Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. *Bayesian Anal.*, **8**, 479 – 504.

Sun, J., Gai, Y. and Lin, L. (2013) Weighted local linear composite quantile estimation for the case of general error distributions. *J. Stat. Plan. Inference*, **143**, 1049–1063.

Tian, Y., Lian, H. and Tian, M. (2017) Bayesian composite quantile regression for linear mixed effects models. *Commun. Stat. - Theory Methods*, **46**, 7717–7731.

Tian, Y., Wang, L., Tang, M. and Tian, M. (2021) Weighted composite quantile regression for longitudinal mixed effects models with application to aids studies. *Commun. Stat. - Simul. Comput.*, **50**, 1837–1853.

Tuithof, M., Ten Have, M., Van Den Brink, W., Vollebergh, W. and De Graaf, R. (2016) Treatment seeking for alcohol use disorders: treatment gap or adequate self-selection? *Eur. Addict. Res.*, **22**, 277–285.

UnitedHealth Group (2019) The high cost of avoidable hospital emergency department visits. UnitedHealth Group, Minnesota (Available from `https://www.unitedhealthgroup.com/newsroom/posts/2019-07-22-high-cost-emergency-department-visits.html`).

U.S. Centers for Medicare & Medicaid (2020) NHE summary, including share of GDP, CY 1960-2019. U.S. Centers for Medicare & Medicaid, Maryland (Available from `https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical`).

Van Rijsbergen, C. J. (1979) *Information Retrieval.* London: Butterworth.

Walker, K., Jiarpakdee, J., Loupis, A., Tantithamthavorn, C., Joe, K., Ben-Meir, M., Akhlaghi, H., Hutton, J., Wang, W., Stephenson, M., Blecher, G., Paul, B., Sweeny, A. and Turhan, B. (2022) Emergency medicine patient wait time multivariable prediction models: a multi-centre derivation and validation study. *Emergency Medicine Journal*, **39**, 386–393.

Wei, J., Defries, T., Lozada, M., Young, N., Huen, W. and Tulsky, J. (2015) An inpatient treatment and discharge planning protocol for alcohol dependence: efficacy in reducing 30-day readmissions and emergency department visits. *J. Gen. Intern. Med.*, **30**, 365–370.

Wei, J. W., Heeley, E. L., Jan, S., Huang, Y., Huang, Q., Wang, J.-G., Cheng, Y., Xu, E., Yang, Q. and Anderson, C. S. (2010) Variations and determinants of hospital costs for acute stroke in china. *PloS One*, **5**, e13041.

Weinstein, Z. M., Wakeman, S. E. and Nolan, S. (2018) Inpatient addiction consult service: Expertise for hospitalized patients with complex addiction problems. *Med. Clin. North Am.*, **102**.

Wennberg, J. and Gittelsohn, A. (1973) Small area variations in health care delivery. *Science*, **182**, 1102–1108.

Witkiewitz, K., Litten, R. Z. and Leggio, L. (2019) Advances in the science and treatment of alcohol use disorder. *Sci. Adv.*, **5**, eaax4043.

Xu, K. (2017) Model-free feature screening via a modified composite quantile correlation. *J. Stat. Plan. Inference*, **188**, 22–35.

Yang, C., Delcher, C., Shenkman, E. and Ranka, S. (2019) Expenditure variations analysis using residuals for identifying high health care utilizers in a state medicaid program. *BMC Med. Inform. Decis. Mak.*, **19**, 131.

Zhao, W.-h., Zhang, R.-q., Lü, Y.-z. and Liu, J.-c. (2016) Bayesian regularized regression based on composite quantile method. *Acta Math. Appl. Sin. Engl. Ser.*, **32**, 495—512.

Zou, H. and Yuan, M. (2008) Composite quantile regression and the oracle model selection theory. *Ann. Stat.*, **36**, 1108–1126.