Bayesian Inference for Spatial Stochastic Volatility Models*

Süleyman Taşpınar† Osman Doğan‡ Jiyoung Chae§ Anil K. Bera¶

November 12, 2019

Abstract

In this study, we propose a spatial stochastic volatility model in which the latent log-volatility terms follow a spatial autoregressive process. Though there is no spatial correlation in the outcome equation (the mean equation), the spatial autoregressive process defined for the log-volatility terms introduces spatial dependence in the outcome equation. To introduce a Bayesian Markov chain Monte Carlo (MCMC) estimation algorithm, we transform the model so that the outcome equation takes the form of log-squared terms. We approximate the distribution of the log-squared error terms in the outcome equation with a finite mixture of normal distributions so that the transformed model turns into a linear Gaussian state-space model. Our simulation results indicate that the Bayesian estimator has satisfactory finite sample properties. We investigate the practical usefulness of our proposed model and estimation method by using the price returns of residential properties in the broader Chicago Metropolitan area.

JEL-Classification: C11, C21, C22.
Keywords: Spatial stochastic volatility, stochastic volatility, SAR model, spatial dependence, Bayesian inference, MCMC, Bayes factor, house price returns.

*This research was supported, in part, by a grant of computer time from the City University of New York High Performance Computing Center under NSF Grants CNS-0855217 and CNS-0958379.

†Department of Economics, Queens College CUNY, New York, U.S., email: staspinar@qc.cuny.edu.
‡Department of Economics, University of Illinois at Urbana-Champaign, U.S., email: odogan@illinois.edu.
§Department of Economics, University of Illinois at Urbana-Champaign, U.S., email: jchae3@illinois.edu.
¶Department of Economics, University of Illinois at Urbana-Champaign, U.S., email: abera@illinois.edu.

1 Introduction

Stochastic volatility models and autoregressive conditional heteroskedasticity (ARCH)-type models are designed to capture volatility clustering phenomenon observed in financial time series. Unlike the ARCH type models, the standard stochastic volatility model consists of separate independent error processes for the conditional mean and the conditional variance. The process for the conditional variance is specified as a log-normal autoregressive process with independent innovations. There is evidence that the stochastic volatility models can offer increased flexibility over the ARCH-type models (Fridman and Harris, 1998; Jacquier et al., 1994, 2004; Kim et al., 1998). The purpose of this paper is to extend the standard stochastic volatility model to spatial data. We suggest a parsimonious specification in which we directly model the log-volatility terms through a first-order spatial autoregressive process. The resulting spatial stochastic volatility model shares similar properties with the standard stochastic volatility model in time series, and it is designed to capture volatility clustering observed in spatial data.

As in the standard stochastic volatility model, our specification consists of two independent error terms for the outcome and the log-volatility equations, respectively. Under the assumption that both error terms have conditional normal distribution, our spatial process implies a leptokurtic symmetric distribution for spatial data. More importantly, though our specification implies no spatial correlation in the outcome variable, the spatial autoregressive process defined for the log-volatility introduces spatial correlation in higher moments of the outcome variable, implying spatial dependence in the outcome variable. To test the spatial dependence in the outcome variable, we formulate a test based on the spatial autoregressive parameter in the log-volatility equation.

To introduce an estimation approach for our specification, we transform the model so that the outcome equation takes the form of log-squared terms. For a similar spatial model, Robinson (2009) approximates the distribution of the log-squared error terms with the normal distribution and establishes the asymptotic consistency and normality of the resulting Gaussian pseudo-maximum likelihood estimator (PMLE). In the time series literature, the PMLEs obtained in this way for the stochastic volatility models also attain the standard large sample properties, but they are sub-optimal in the sense that they have poor finite sample properties (Jacquier et al., 1994; Kim et al., 1998; Sandmann and Koopman, 1998; Shephard, 1994). Therefore, we propose approximating the distribution of the log-squared error terms by a mixture of Gaussian distributions (Kim et al., 1998; Shephard, 1994). The resulting estimation system turns into a Gaussian state-space model, where the log-volatility equation constitutes the state equation. We then introduce a Bayesian MCMC estimation approach in which a data augmentation scheme is used to treat the latent log-volatility terms as additional parameters, which are estimated as a natural by-product of the estimation process. In a Monte Carlo study, we investigate the finite sample properties of our Bayesian estimator along with a (naive) Bayesian estimator based on an algorithm in which the distribution of log-squared error terms is approximated by the normal distribution. Our results indicate that the naive estimator has poor finite sample properties, whereas the Bayesian estimator based on the finite mixture of normal distributions performs well. To test the presence of spatial

correlation in the log-volatility, i.e., to test spatial dependence in the outcome variable, we suggest using the Savage-Dickey density ratio (SDDR) for the calculation of the Bayes factor (Dickey, 1971; Verdinelli and Wasserman, 1995).

To investigate the practical usefulness of our proposed methodology, we apply our specification to the price returns of the residential properties in the broader Chicago Metropolitan area for the years of 2014 and 2015. We first use the Moran I test to check for the presence of spatial correlations in the returns and squared returns. The results indicate that there is a mild indication for spatial correlation in the return series; however, a very strong evidence for spatial correlation in the squared returns. Based on these results, the housing market may not be efficient and therefore we estimate a specification that allows for spatial correlations both in the returns and the log-volatility terms. We show that the estimated spatial autoregressive parameters are significant. The conditional variance estimates indicate that the lower estimates are scattered over the suburbs of the city of Chicago, while relatively larger estimates are concentrated over a corridor extending from the west side of the city to the south side.

The rest of the paper is organized as follows. In Section 2, we provide a brief review of the related literature, where we first describe the theory related to our estimation approach, and then the literature on house price formations. We discuss the empirical justification for specification based on the stylized facts in the context of price returns in the residential properties, in particular, in the city of Chicago and its surrounding suburbs. In Section 3, we state our model specification and discuss its properties. In Section 4, we show how the Bayesian MCMC method can be used to estimate our model. In Section 5, we develop a test based on the SDDR to test the presence of spatial correlation in log-volatility terms. In Section 6, we consider some extension of our parsimonious model and show how the Bayesian estimation approach should be adjusted accordingly. In Section 7, we investigate the finite sample properties of our suggested algorithms through a Monte Carlo study. In Section 8, we provide the details of our empirical application on house price returns. In Section 9, we offer concluding comments.

2 Related Literature

The model specification suggested in this paper belongs to a recent and growing literature on spatial econometric models. These models extend the conventional regression models by including spatial lag(s) of the dependent variable (and/or spatial lag(s) of the disturbance terms). Spatial models formulated in this way are called spatial autoregressive models, which can be considered as the empirical counterparts for the equilibrium outcome of theoretical models of interacting spatial units (Anselin, 1988, 2007; Baltagi et al., 2014; Cliff and Ord, 1972; Doğan and Taşpınar, 2018; Elhorst, 2014; Kelejian and Prucha, 2010; Lee, 2004, 2007; Lee et al., 2010; LeSage and Pace, 2009; Ord, 1975). Recently, some studies have also used the terminology, *weak* and *strong* spatial dependence, to respectively refer to regression models that have spatial lag terms and interactive fixed effects (Bailey et al., 2016; Chudik et al., 2011; Han and Lee, 2016; Pesaran, 2015; Shi and

Lee, 2017). Robinson (2009) introduces a spatial process that allows for no spatial correlation in the dependent variable, but spatial correlation in the higher moment of the dependent variable. Our paper is closely related to Robinson (2009), but differs in two important aspects. First, we consider extended versions that allow for spatial correlation in the dependent variable and in the higher moment of the dependent variable. Second, instead of using the PMLE suggested in Robinson (2009), we propose the Bayesian MCMC estimation approach with a data augmentation scheme for estimation. In our approach, the estimates of the latent log-volatility terms are produced as a natural by-product of the estimation process.

In an empirical application, we focus on the spatial dependence in the log-volatilities of the returns calculated from house prices in the broader Chicago Metropolitan area. Spatial correlation in house price variations may arise due to several factors. For example, Meen (1999) points to migration, equity transfer, spatial arbitrage and spatial patterns in the determinants of house prices. Bailey et al. (2016) use a two-stage approach to analyze strong and weak spatial dependence in house price changes at the level of Metropolitan Statistical Areas (MSA) in the US over the period 1975 to 2010. Their results show statistically significant positive and negative spill-over effects across MSAs in the USA over the period 1975 to 2010, with the positive effects being more prevalent. There are also studies incorporating spatial correlation into the standard vector autoregression models (VARs) to explore spatio-temporal diffusion of house prices through impulse response analysis (Beenstock and Felsenstein, 2007; Brady, 2011; Holly et al., 2011; Kuethe and Pede, 2011). A significant bulk of literature focuses on testing the cointegration relationships between house prices and their determinants implied by the arbitrage equations derived from theoretical models (Gallin, 2006; Holly et al., 2010; Meen, 1996). Holly et al. (2010) consider a model in which the price of a house is determined by setting the expected net benefit from owning the house against the real rental cost of the same property. The market equilibrium implies a cointegrating relationship between the real price of housing and the real per capita personal disposable income, and Holly et al. (2010) use annual US states level data from 1975 to 2003 to establish panel cointegration by taking explicit account of both cross-sectional dependence and parameter heterogeneity.

In this paper, we focus on the returns calculated from the first difference of log house prices and suggest a parsimonious formulation to model spatial dependence in returns. Our formulation allows us to make a clear distinction between spatial correlation (linear dependence) and spatial dependence (non-linear dependence) in the returns. Our approach allows for modeling conditional mean and conditional variance of returns through separate spatial autoregressive processes. Furthermore, our Bayesian estimation approach yields estimates of latent conditional variance terms. For the empirical justification of our formulation, we use the residential property sale prices in the city of Chicago and its surrounding suburbs for years 2014 and 2015. As shown in the scatter plots of the returns and squared returns in Figure 1, the spatial clustering in returns is somewhat weak, however, there are spatial clustering patterns in squared returns, suggesting that the conditional variance of the returns may vary over space.[1] Indeed, as we will show in Section 8, the Moran I

[1]The returns and squared returns are classified into low and high categories according to whether they are smaller

test rejects the null hypothesis of no spatial correlation in the squared returns.

3 Model Specification

Following the time series literature on stochastic volatility models, we specify the following outcome equation

$$y_i = e^{\frac{1}{2}h_i}\,\varepsilon_i, \quad \text{for} \quad i = 1, \ldots, n, \tag{3.1}$$

where ε_i is an independent and identically distributed (i.i.d) normal random variable with mean zero and unit variance for $i = 1, \ldots, n$. We assume a first-order spatial autoregressive process for h_i:

$$h_i - \mu_h = \lambda \sum_{j=1}^{n} w_{ij}(h_j - \mu_h) + u_i, \tag{3.2}$$

where μ_h is the constant mean, w_{ij}'s for $i, j = 1, \ldots, n$ are non-stochastic spatial weights, and u_i is an i.i.d normal random variable with mean zero and variance σ_u^2. We assume that u_i and ε_j are independent for all i, j.[2] The scalar parameter λ is the spatial autoregressive parameter and measures the degree of spatial correlation among h_i's. It follows that the conditional variance of y_i given h_i is

$$\text{Var}\,(y_i|h_i) = e^{h_i}, \tag{3.3}$$

which makes the conditional variance of y_i is *space-varying*. Thus, analogous to time-series literature on stochastic volatility models, we refer to h_i as the *log-volatility*. Let $\mathbf{h} = (h_1, \ldots, h_n)'$ be the $n \times 1$ vector of log-volatilities and $\mathbf{u} = (u_1, \ldots, u_n)'$ be the $n \times 1$ vector of disturbance terms. Then, (3.2) can be written in vector form as

$$\mathbf{h} - \boldsymbol{l}_n\mu_h = \lambda \mathbf{W}(\mathbf{h} - \boldsymbol{l}_n\mu_h) + \mathbf{u}, \tag{3.4}$$

where $\mathbf{W} = (w_{ij})$ is the $n \times n$ non-stochastic spatial weights matrix that has zero diagonal elements and $\boldsymbol{l}_n$ is the $n \times 1$ vector of ones. The spatial autoregressive model is an equilibrium model, hence we assume that $\mathbf{h} - \boldsymbol{l}_n\mu_h = \mathbf{S}^{-1}(\lambda)\mathbf{u}$ exists, where $\mathbf{S}(\lambda) = (\mathbf{I}_n - \lambda\mathbf{W})$ and $\mathbf{I}_n$ is the $n \times n$ identity matrix.

The process y_i defined through (3.1) and (3.2) exhibits no spatial correlation since $\mathbb{E}(y_i y_j) = 0$ for all $i \neq j$. Furthermore, all odd moments of y_i are zero since ε_i has the standard normal distribution and is independent of h_i for all i. Let $\mathbf{K}_i(\lambda)$ be the ith row vector of $\mathbf{S}^{-1}(\lambda)$ and $r \in \mathbb{N}$

or larger than their corresponding sample averages. The details of this data set are provided in Section 8.

[2]Our parsimonious model consists of (3.1) and (3.2). In Section 6, we consider some extensions and show how the Bayesian estimation should be adjusted accordingly.

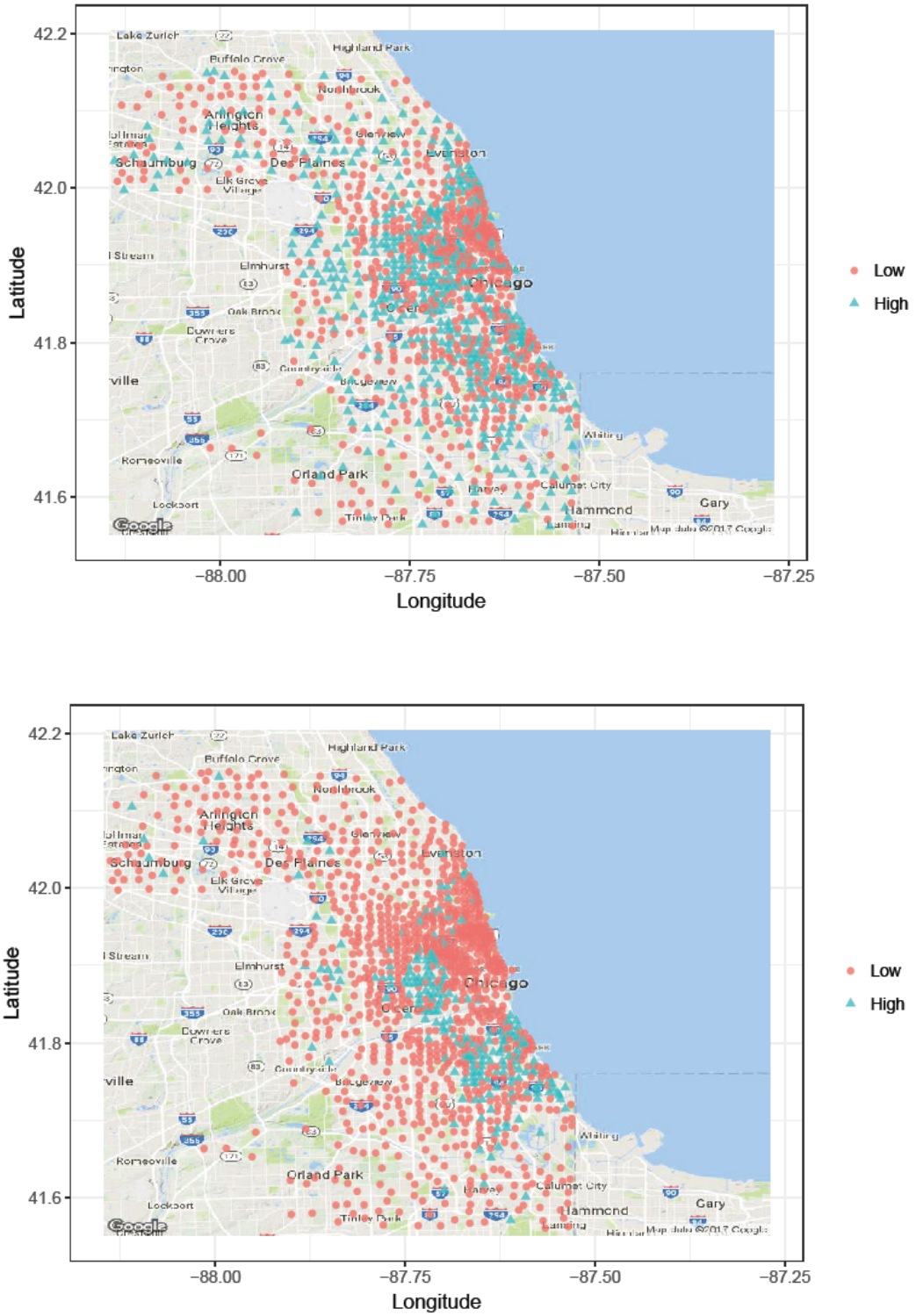


Figure 1: Scatter plots: (a) Returns (top), (b) squared returns (bottom)

be an even number. Then, all even moments of y_i exist and are given by

$$\mathbb{E}(y_i^r) = \mathbb{E}\left(e^{\frac{1}{2}h_i r}\right)\mathbb{E}\left(\varepsilon_i^r\right) = e^{\frac{\mu_h r}{2}+\frac{\sigma_u^2 r^2}{8}\|\mathbf{K}_i(\lambda)\|^2}\mu_{(r)}, \tag{3.5}$$

where $\mu_{(r)} = \frac{r!}{2^{r/2}\times(r/2)!}$. Using (3.5), it can be shown that $\mathbb{E}(y_i^4)/\left[\mathbb{E}(y_i^2)\right]^2 - 3 = 3\left(e^{\sigma_u^2\|\mathbf{K}_i(\lambda)\|^2} - 1\right) > 0$. Thus, y_i has a leptokurtic symmetric distribution. Now, consider $y_i^r = e^{\frac{r}{2}h_i}\,\varepsilon_i^r$ for $r \in \mathbb{N}$ even. It can be shown that

$$\begin{aligned}
\mathrm{Cov}(y_i^r,\, y_j^r) &= \mathbb{E}(e^{\frac{r}{2}(h_i+h_j)})\mathbb{E}(\varepsilon_i^r)\mathbb{E}(\varepsilon_j^r) - \mathbb{E}(e^{\frac{r}{2}h_i})\mathbb{E}(\varepsilon_i^r)\mathbb{E}(e^{\frac{r}{2}h_j})\mathbb{E}(\varepsilon_j^r) \\
&= \mu_{(r)}^2 e^{\mu_h r+\frac{\sigma_u^2 r^2}{8}\left(\|\mathbf{K}_i(\lambda)\|^2+\|\mathbf{K}_j(\lambda)\|^2+2\mathbf{K}_i'(\lambda)\mathbf{K}_j(\lambda)\right)} \\
&\quad - \mu_{(r)}^2 e^{\frac{\mu_h r}{2}+\frac{\sigma_u^2 r^2}{8}\|\mathbf{K}_i(\lambda)\|^2} e^{\frac{\mu_h r}{2}+\frac{\sigma_u^2 r^2}{8}\|\mathbf{K}_j(\lambda)\|^2} \\
&= \mu_{(r)}^2\left[e^{\mu_h r+\frac{\sigma_u^2 r^2}{8}\left(\|\mathbf{K}_i(\lambda)\|^2+\|\mathbf{K}_j(\lambda)\|^2\right)}\left(e^{\frac{\sigma_u^2 r^2}{8}2\mathbf{K}_i'(\lambda)\mathbf{K}_j(\lambda)} - 1\right)\right].
\end{aligned} \tag{3.6}$$

The covariance in (3.6) is generally not zero, implying spatial dependence for y_i's. When $\lambda = 0$, we obtain $\mathrm{Cov}(y_i^r,\, y_j^r) = 0$, because $\left(e^{\frac{\sigma_u^2 r^2}{8}2\mathbf{K}_i'(\lambda)\mathbf{K}_j(\lambda)} - 1\right) = 0$. Thus, the presence of spatial dependence in y_i's can be tested by a test statistic for the null hypothesis of $H_0 : \lambda = 0$.[3]

For the Bayesian analysis, we transform the model so that the resulting estimation equation becomes linear in the log-volatility h_i. Thus, we square both sides of (3.1) and then take the natural logarithm to obtain

$$y_i^* = h_i + \varepsilon_i^*, \tag{3.7}$$

where $y_i^* = \log y_i^2$, and $\varepsilon_i^* = \log \varepsilon_i^2$. Note that ε_i^* has a $\log \chi_1^2$ distribution with the density

$$p(\varepsilon_i^*) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(e^{\varepsilon_i^*} - \varepsilon_i^*)\right), \quad -\infty < \varepsilon_i^* < \infty, \quad i = 1, 2, \ldots, n. \tag{3.8}$$

It can be shown that $\mathbb{E}(\varepsilon_i^*) \approx -1.2704$ and $\mathrm{Var}(\varepsilon_i^*) = \pi^2/2 \approx 4.9348$. The density given in (3.8) is highly skewed with a long left tail.

Let $\mathbf{y}^* = (y_1^*, \ldots, y_n^*)'$ be the $n \times 1$ vector of the transformed dependent variable, and $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, \ldots, \varepsilon_n^*)'$ be the $n \times 1$ vector of transformed disturbance terms. Then, in vector form, we have

$$\mathbf{y}^* = \mathbf{h} + \boldsymbol{\varepsilon}^*. \tag{3.9}$$

Analogous to time series literature, we note that (3.9) and (3.4) define a linear state space model in $\mathbf{h}$. In the time series setting, the state equation determines how the state variable is generated from the time-lags of the state variable. In our case instead, the state variable $\mathbf{h}$ depends on its *spatial-lag* term $\mathbf{Wh}$ as shown in (3.4).

[3] In Section 4, we show how this null hypothesis can be tested.

Table 1: The ten-component Gaussian mixture for $\log \chi_1^2$

Components	p_j	μ_j	σ_j^2
1	0.00609	1.92677	0.11265
2	0.04775	1.34744	0.17788
3	0.13057	0.73504	0.26768
4	0.20674	0.02266	0.40611
5	0.22715	-0.85173	0.62699
6	0.18842	-1.97278	0.98583
7	0.12047	-3.46788	1.57469
8	0.05591	-5.55246	2.54498
9	0.01575	-8.68384	4.16591
10	0.00115	-14.65000	7.33342

4 Posterior Analysis

In this section, we determine the conditional posterior distributions of the model parameters. We approximate the density function $p(\varepsilon_i^*)$ given in (3.8) using a mixture of Gaussian distributions (Kim et al., 1998; Shephard, 1994). More precisely, we consider the following m-component Gaussian mixture distribution:

$$p(\varepsilon_i^*) \approx \sum_{j=1}^{m} p_j \times \phi(\varepsilon_i^* \,|\mu_j,\, \sigma_j^2), \tag{4.1}$$

where $\phi(\varepsilon_i^* \,|\mu_j,\, \sigma_j^2)$ denotes the Gaussian density function with mean μ_j and variance σ_j^2, p_j is the probability of jth mixture component and m is the number of components. We can equivalently represent (4.1) in terms of an auxiliary random variable $s_i \in \{1, 2, \ldots, m\}$ that serves as the mixture component indicator:

$$\varepsilon_i^* \,|(s_i = j) \sim N(\mu_j,\, \sigma_j^2), \quad \text{and} \quad \mathbb{P}(s_i = j) = p_j, \quad j = 1, 2, \ldots, m, \quad i = 1, 2, \ldots, n. \tag{4.2}$$

Thus, the model in (3.7) is now conditionally linear and Gaussian given the component indicator variable $\mathbf{s} = (s_1, \ldots, s_n)$. We are now in a position to determine the elements of (4.1) to make the mixture approximation sufficiently accurate. Following Omori et al. (2007), we use a ten-component Gaussian mixture distribution whose parameters are given in Table 1. Note that these parameter values do not vary during estimation, therefore the mixture Gaussian approximation approach does not induce any additional computation time.

To complete the model specification, we assume the following independent prior distributions for σ_u^2, λ and μ_h: $\sigma_u^2 \sim IG(a_0, b_0)$, $\lambda \sim$ Uniform$(-1/\tau,\, 1/\tau)$, $\mu_h \sim N(\mu_0, V_\mu)$, where IG$(a_0,\, b_0)$ is the inverse gamma distribution with shape parameter a_0 and scale parameter b_0, and Uniform$(-1/\tau,\, 1/\tau)$ is the uniform distribution, where τ is the spectral radius of $\mathbf{W}$. A closed subset of the interval $(-1/\tau,\, 1/\tau)$ can be considered as the parameter space for λ since $\mathbf{S}(\lambda)$ is invertible for all

$\lambda \in (-1/\tau,\, 1/\tau)$.[4] Using the Bayes' theorem, the joint posterior distribution $p(\mathbf{h}, \mathbf{s}, \sigma_u^2, \mu_h, \lambda|\mathbf{y}^*)$ can be stated as

$$p(\mathbf{h}, \mathbf{s}, \sigma_u^2, \mu_h, \lambda|\mathbf{y}^*) \propto p(\mathbf{y}^*|\mathbf{h}, \mathbf{s}) \times p(\mathbf{h}|\sigma_u^2, \mu_h, \lambda) \times p(\mathbf{s}) \times p(\sigma_u^2) \times p(\mu_h) \times p(\lambda), \tag{4.3}$$

where $p(\mathbf{y}^*|\mathbf{h}, \mathbf{s})$ is the conditional likelihood function of (3.7). Let $\mathbf{d}_s = (\mu_{s_1}, \ldots, \mu_{s_n})'$ and $\mathbf{\Sigma}_s = \text{Diag}(\sigma_{s_1}^2, \ldots, \sigma_{s_n}^2)$. Then, from (4.2), we have $\boldsymbol{\varepsilon}^*|\mathbf{s} \sim N(\mathbf{d}_s,\, \mathbf{\Sigma}_s)$. It follows from (3.7) and (3.2) respectively that

$$\mathbf{y}^*|\mathbf{s},\, \mathbf{h} \sim N\left(\mathbf{h} + \mathbf{d}_s,\, \mathbf{\Sigma}_s\right), \tag{4.4}$$

$$\mathbf{h}|\sigma_u^2, \mu_h, \lambda \sim N\left(\boldsymbol{l}_n\mu_h,\, \sigma_u^2\mathbf{S}^{-1}(\lambda)\mathbf{S}'^{-1}(\lambda)\right). \tag{4.5}$$

We consider the following Gibbs sampler to generate random draws from $p(\mathbf{h}, \mathbf{s}, \sigma_u^2, \mu_h, \lambda|\mathbf{y}^*)$.

Algorithm 1 (Estimation algorithm based on the Gaussian mixture distribution)**.**

1. Sampling step for $\mathbf{s}$:

$$\mathbb{P}(s_i = j|y_i^*, h_i) = \frac{p_j \times \phi\left(y_i^*|h_i + \mu_j,\, \sigma_j^2\right)}{\sum_{k=1}^{10} p_k \times \phi\left(y_i^*|h_i + \mu_k,\, \sigma_k^2\right)}, \quad j = 1, \ldots, 10, \quad i = 1, \ldots, n, \tag{4.6}$$

   where the denominator is the normalizing constant.

2. Sampling step for $\mathbf{h}$:

$$\mathbf{h}|\mathbf{y}^*, \mathbf{s}, \sigma_u^2, \mu_h, \lambda \sim N\left(\widehat{\mathbf{h}},\, \widehat{\mathbf{H}}_h^{-1}\right), \tag{4.7}$$

   where $\widehat{\mathbf{H}}_h = \mathbf{\Sigma}_s^{-1} + \frac{1}{\sigma_u^2}\mathbf{S}'(\lambda)\mathbf{S}(\lambda)$ and $\widehat{\mathbf{h}} = \widehat{\mathbf{H}}_h^{-1}\left(\mathbf{\Sigma}_s^{-1}(\mathbf{y}^* - \mathbf{d}_s) + \frac{\mu_h}{\sigma_u^2}\mathbf{S}'(\lambda)\mathbf{S}(\lambda)\boldsymbol{l}_n\right)$.

3. Sampling step for σ_u^2:

$$\sigma_u^2|\mathbf{h}, \mu_h, \lambda \sim \text{IG}\left(a_0 + \frac{n}{2},\, b_0 + \frac{1}{2}(\mathbf{h} - \boldsymbol{l}_n\mu_h)'\mathbf{S}'(\lambda)\mathbf{S}(\lambda)(\mathbf{h} - \boldsymbol{l}_n\mu_h)\right). \tag{4.8}$$

4. Sampling step for μ_h:

$$\mu_h|\mathbf{h}, \sigma_u^2, \lambda \sim N\left(\widehat{\mu}_0, \widehat{V}_\mu^{-1}\right), \tag{4.9}$$

   where $\widehat{V}_\mu = V_\mu^{-1} + \frac{1}{\sigma_u^2}\boldsymbol{l}_n'\mathbf{S}'(\lambda)\mathbf{S}(\lambda)\boldsymbol{l}_n$ and $\widehat{\mu}_0 = \widehat{V}_\mu^{-1}\left(\frac{1}{\sigma_u^2}\boldsymbol{l}_n'\mathbf{S}'(\lambda)\mathbf{S}(\lambda)\mathbf{h} + V_\mu^{-1}\mu_0\right)$.

[4] See LeSage and Pace (2009) and Kelejian and Prucha (2010) for a discussion on the parameter space for spatial autoregressive parameters.

5. Sampling step for λ:

$$p(\lambda|\mathbf{h},\mu_h,\sigma_u^2) \propto |\mathbf{S}(\lambda)| \times \exp\left(-\frac{1}{2\sigma_u^2}(\mathbf{h}-\boldsymbol{l}_n\mu_h)'\mathbf{S}'(\lambda)\mathbf{S}(\lambda)(\mathbf{h}-\boldsymbol{l}_n\mu_h)\right), \tag{4.10}$$

which does not correspond to any known density function. A random-walk Metropolis-Hastings algorithm can be used to sample from this distribution (LeSage and Pace, 2009). A candidate value λ^{new} is generated according to

$$\lambda^{new} = \lambda^{old} + z_\lambda \times N(0,1), \tag{4.11}$$

where z_λ is the tuning parameter. The candidate value λ^{new} is accepted with probability $\mathbb{P}(\lambda^{new}, \lambda^{old}) = \min\left(1, \frac{p(\lambda^{new}|\mathbf{h},\sigma_u^2,\mu_h)}{p(\lambda^{old}|\mathbf{h},\sigma_u^2,\mu_h)}\right)$.

Remark 1. *In sampling* $\mathbf{h}$*, we can reduce the computational burden by first obtaining the Cholesky factor* $\mathbf{C}$ *such that* $\mathbf{C}'\mathbf{C} = \widehat{\mathbf{H}}_h$*, and then solving* $\mathbf{C}'\mathbf{C}\widehat{\mathbf{h}} = \boldsymbol{\Sigma}_s^{-1}(\mathbf{y}^* - \mathbf{d}_s) + \frac{\mu_h}{\sigma_u^2}\mathbf{S}'(\lambda)\mathbf{S}(\lambda)\boldsymbol{l}_n$ *for* $\widehat{\mathbf{h}}$*. Then, we can sample from* $N\left(\widehat{\mathbf{h}}, \widehat{\mathbf{H}}_h^{-1}\right)$ *by first drawing from* $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I}_n)$ *and solving* $\mathbf{C}\boldsymbol{\xi} = \mathbf{u}$ *for* $\boldsymbol{\xi}$*, and finally setting* $\mathbf{h} = \widehat{\mathbf{h}} + \boldsymbol{\xi}$ *(Chan and Jeliazkov, 2009).*

Remark 2. *In Step 5, the tuning parameter* z_λ *is adjusted such that the acceptance rate falls between* 40% *and* 60% *during the sampling process (LeSage and Pace, 2009). In sampling* λ*, we can alternatively consider an independence-chain Metropolis-Hastings algorithm or a Griddy-Gibbs algorithm. In the case of independence-chain Metropolis-Hastings algorithm, a tailored normal distribution* $N(\mu_\lambda, \sigma_\lambda^2)$*, where* μ_λ *is the mode of* $\log p(\lambda|\mathbf{h},\sigma_u^2,\mu_h)$ *and* σ_λ^2 *is the inverse of the negative Hessian of* $\log p(\lambda|\mathbf{h},\sigma_u^2,\mu_h)$ *evaluated at the mode, can be used to generate candidate values (Chib and Greenberg, 1994, 1995). The mode of* $\log p(\lambda|\mathbf{h},\sigma_u^2,\mu_h)$ *can be found by using the Newton-Raphson recursion based on*

$$\frac{\partial \log p(\lambda|\mathbf{h},\sigma_u^2,\mu_h)}{\partial\lambda} = -\mathrm{tr}\left(\mathbf{S}^{-1}(\lambda)\mathbf{W}\right) + \frac{1}{\sigma_u^2}(\mathbf{h}-\boldsymbol{l}_n\mu_h)'\mathbf{W}'\mathbf{S}(\lambda)(\mathbf{h}-\boldsymbol{l}_n\mu_h), \tag{4.12}$$

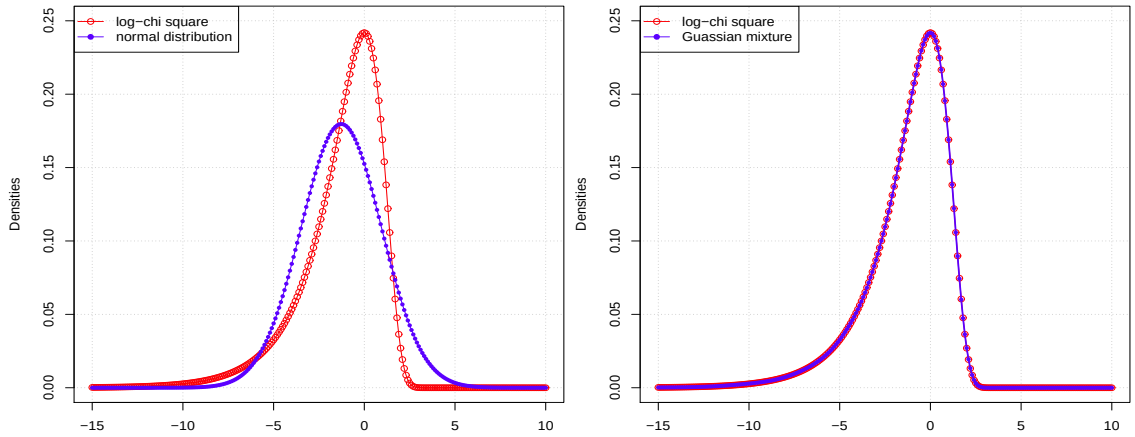
$$\frac{\partial^2 \log p(\lambda|\mathbf{h},\sigma_u^2,\mu_h)}{\partial\lambda^2} = -\mathrm{tr}\left([\mathbf{S}^{-1}(\lambda)\mathbf{W}]^2\right) - \frac{1}{\sigma_u^2}(\mathbf{h}-\boldsymbol{l}_n\mu_h)'\mathbf{W}'\mathbf{W}(\mathbf{h}-\boldsymbol{l}_n\mu_h). \tag{4.13}$$

Then, the candidate value λ^{new} *generated from* $N(\mu_\lambda, \sigma_\lambda^2)$ *is accepted with probability* $P(\lambda^{new}, \lambda^{old}) = \min\left(1, \frac{p(\lambda^{new}|\mathbf{h},\sigma_u^2,\mu_h)\times\phi(\lambda^{old}|\mu_\lambda,\sigma_\lambda^2)}{p(\lambda^{old}|\mathbf{h},\sigma_u^2,\mu_h)\times\phi(\lambda^{new}|\mu_\lambda,\sigma_\lambda^2)}\right)$.

The Griddy-Gibbs sampler can be considered as a discretized version of the inverse-transform method and it only requires the evaluation of the target density (Ritter and Tanner, 1992). Here, the conditional posterior distribution of λ *is approximated by a discretized distribution on a fine grid formed from the parameter space of* λ*. Note that, in terms of computational efficiency, the random-walk Metropolis-Hastings algorithm is relatively more efficient than the other two algorithms. Therefore, we use this algorithm to generate draws for* λ *in our simulation.*

An alternative approach for the estimation of our model can be based on the approximation of the normal distribution to the distribution of $\log\varepsilon_i^2$ (Harvey et al., 1994; Robinson, 2009; Ruiz,

1994). Although the QMLE obtained based on this approximation method has the standard large sample properties, it is sub-optimal in the sense that it has poor finite sample properties because the distribution of $\log \varepsilon_i^2$ is poorly approximated by the normal distribution. We compare the ten-component Gaussian mixture (given in Table 1) and the standard normal distribution in terms of how well they approximate to the distribution of $\log \varepsilon_i^2$. In Figure 2, we plot the density of each distribution. Figure 2(a) shows the densities of $\log \chi_1^2$ and the Gaussian distribution that has a mean of -1.2704 and a variance of $\pi^2/2$. In Figure 2(b), we plot the densities of $\log \chi_1^2$ and the ten-component Gaussian mixture distribution. The ten-component Gaussian mixture distribution approximates the distribution of $\log \varepsilon_i^2$ very closely.



(a) Densities of $\log \chi_1^2$ and $N(-1.2704, \pi^2/2)$ (b) The ten-component Gaussian-mixture and $\log \chi_1^2$ densities

Figure 2: Densities of $\log \chi_1^2$ and its approximations

Next, we illustrate the Bayesian estimation under the Gaussian approximation. Let $\xi_i = \log \varepsilon_i^2 - \mathbb{E}(\log \varepsilon_i^2)$. Then, we can express (3.7) in terms of ξ_i as

$$y_i^\circ = h_i + \xi_i, \tag{4.14}$$

where $y_i^\circ = y_i^* + 1.2704$. Let $\mathbf{y}^\circ = (y_i^\circ, \ldots, y_n^\circ)'$ be the $n \times 1$ vector. Under the assumption that ξ_i is i.i.d Gaussian with mean zero and variance $\pi^2/2$, we have

$$\mathbf{y}^\circ | \mathbf{h} \sim N\left(\mathbf{h},\, \pi^2/2\mathbf{I}_n\right). \tag{4.15}$$

Given our assumed prior distributions for σ_u^2, μ_h and λ, the following four steps Gibbs sampler can be considered to generate random draws from $p(\mathbf{h}, \sigma_u^2, \mu_h, \lambda | \mathbf{y}^\circ)$.

Algorithm 2 (Estimation algorithm based on Gaussian approximation)**.**

1. Sampling step for $\mathbf{h}$:

$$h|y^{\circ}, \sigma_u^2, \mu_h, \lambda \sim N\left(\mathbf{h}^{\dagger}, \mathbf{H}_h^{\dagger -1}\right), \tag{4.16}$$

where $\mathbf{H}_h^{\dagger} = \frac{2}{\pi^2}\mathbf{I}_n + \frac{1}{\sigma_u^2}\mathbf{S}'(\lambda)\mathbf{S}(\lambda)$ and $\mathbf{h}^{\dagger} = \mathbf{H}_h^{\dagger -1}\left(\frac{2}{\pi^2}\mathbf{y}^{\circ} + \frac{\mu_h}{\sigma_u^2}\mathbf{S}'(\lambda)\mathbf{S}(\lambda)\boldsymbol{l}_n\right)$.

2. Sampling step for σ_u^2:

$$\sigma_u^2|\mathbf{h}, \mu_h, \lambda \sim \text{IG}\left(a_0 + \frac{n}{2},\, b_0 + \frac{1}{2}(\mathbf{h} - \boldsymbol{l}_n\mu_h)'\mathbf{S}'(\lambda)\mathbf{S}(\lambda)(\mathbf{h} - \boldsymbol{l}_n\mu_h)\right). \tag{4.17}$$

3. Sampling step for μ_h:

$$\mu_h|\mathbf{h}, \sigma_u^2, \lambda \sim N\left(\mu_0^{\dagger}, V_{\mu}^{\dagger -1}\right), \tag{4.18}$$

where $V_{\mu}^{\dagger} = V_h^{-1} + \frac{1}{\sigma_u^2}\boldsymbol{l}_n'\mathbf{S}'(\lambda)\mathbf{S}(\lambda)\boldsymbol{l}_n$ and $\mu_0^{\dagger} = V_{\mu}^{\dagger -1}\left(V_h^{-1}\mu_0 + \frac{1}{\sigma_u^2}\boldsymbol{l}_n'\mathbf{S}'(\lambda)\mathbf{S}(\lambda)\mathbf{h}\right)$.

4. Sampling step for λ:

$$p(\lambda|\mathbf{h}, \sigma_u^2, \mu_h) \propto |S(\lambda)| \times \exp\left(-\frac{1}{2\sigma_u^2}(\mathbf{h} - \boldsymbol{l}_n\mu_h)'\mathbf{S}'(\lambda)\mathbf{S}(\lambda)(\mathbf{h} - \boldsymbol{l}_n\mu_h).\right), \tag{4.19}$$

which does not correspond to a known form. We can use the random-walk Metropolis-Hastings algorithm described in Step 5 of Algorithm 1 to sample this parameter.

5 Testing Spatial Dependence

In this section, we consider a test for the null hypothesis of no spatial dependence, i.e., $H_0 : \lambda = 0$ against $H_1 : \lambda \neq 0$ in (3.2). In the Bayesian approach, hypothesis testing can be considered as a model comparison exercise and thus can be conducted through the Bayes factors calculated for the competing models (see Kass and Raftery (1995) for a survey on the Bayes factors). Since our null hypothesis requires a nested model comparison, we consider the Savage-Dickey density ratio (SDDR) proposed by Dickey (1971) and Verdinelli and Wasserman (1995) for calculating the Bayes factor. Let M_R and M_U be respectively the restricted and the unrestricted model. Then, the Bayes factor in favor of the unrestricted model is

$$\text{BF}_{UR} = \frac{p(\mathbf{y}^*|M_U)}{p(\mathbf{y}^*|M_R)}, \tag{5.1}$$

where $p(\mathbf{y}^*|M_j) = \int p(\mathbf{y}^*|\boldsymbol{\theta}_j, M_j) \times p(\boldsymbol{\theta}_j|M_j)\text{d}\theta_j$ is the corresponding marginal likelihood or marginal data density for $j \in \{U, R\}$, and $\boldsymbol{\theta}_j$ is the corresponding parameter vector in the competing models. Since our prior distributions are independent, the Bayes factor in (5.1) reduces to the

SDDR given by (Verdinelli and Wasserman, 1995)

$$\text{BF}_{UR} = \frac{p(\lambda = 0|M_U)}{p(\lambda = 0|\mathbf{y}^*, M_U)}, \tag{5.2}$$

where $p(\lambda = 0|M_U)$ and $p(\lambda = 0|\mathbf{y}^*, M_U)$ are respectively the prior and the marginal posterior of λ evaluated at $\lambda = 0$. Thus, if $p(\lambda = 0|M_U)$ is larger than $p(\lambda = 0|\mathbf{y}^*, M_U)$, i.e., if $\lambda = 0$ is more likely under the prior relative to the marginal posterior, then BF_{UR} provides evidence in favor of H_1. Given the Uniform$(-1/\tau, 1/\tau)$ prior on λ, we have $p(\lambda = 0|M_U) = \tau/2$, and the marginal posterior $p(\lambda = 0|\mathbf{y}^*, M_U)$ can be estimated by the following Rao-Blackwell estimator:

$$\widehat{p}(\lambda = 0|\mathbf{y}^*, M_U) = \frac{1}{R}\sum_{r=1}^{R} p(\lambda = 0|\mathbf{y}^*, \mathbf{h}^r, \mathbf{s}^r, \sigma_u^{2r}, \mu_h^r), \tag{5.3}$$

where $\{\mathbf{h}^r, \mathbf{s}^r, \sigma_u^{2r}, \mu_h^r\}_{r=1}^R$ is the sequence of random draws generated through Algorithm 1 or Algorithm 2 (Gelfand and Smith, 1990). Note that the estimator in (5.3) requires that the conditional density of λ is in a standard form, which is not the case given our result in (4.10). However, the conditional posterior density of λ is bounded on the interval $(-1/\tau, 1/\tau)$, and thus we can evaluate the density on a grid given the posterior draws. In the following algorithm, we show how $p(\lambda = 0|\mathbf{y}^*, M_U)$ can be evaluated.

Algorithm 3 (Calculating SDDR)**.**

1. Construct a grid with random points $\lambda_1, \ldots, \lambda_m$ from the interval $(-1/\tau, 1/\tau)$, where τ is the spectral radius of $\mathbf{W}$. The grid must include $\lambda_k = 0$.
2. Compute $p_r(\lambda_i) = \frac{p(\lambda_i|\mathbf{y}^*, \mathbf{h}^r, \mathbf{s}^r, \sigma_u^{2r}, \mu_h^r)}{\sum_{j=1}^m p(\lambda_j|\mathbf{y}^*, \mathbf{h}^r, \mathbf{s}^r, \sigma_u^{2r}, \mu_h^r)}$ for $i = 1, \ldots, m$ and $r = 1, \ldots, R$.
3. Compute $p(\lambda_i) = \sum_{r=1}^R p_r(\lambda_i)$ for $i = 1, \ldots, m$.
4. Then, $\widehat{p}(\lambda = 0|\mathbf{y}^*, M_U) = p(\lambda_k)$.

6 Some Extensions

In this section, we extend out basic model in (3.1) in three directions and show how the suggested Gibbs samplers can be adjusted accordingly. In the first extension, we assume that the process has a deterministic mean equation determined by a vector of exogenous variables. More specifically, we consider

$$y_i = \mathbf{x}_i^{'}\boldsymbol{\beta} + e^{\frac{1}{2}h_i}\,\varepsilon_i, \quad \text{for} \quad i = 1, \ldots, n, \tag{6.1}$$

where $\mathbf{x}_i$ is the $k \times 1$ vector of exogenous variables with the matching parameter vector $\boldsymbol{\beta}$. It is obvious that this model has the same characteristics as our simple model in (3.1), except having a non-zero mean for the outcome variable. Our auxiliary mixture sampler introduced in Section 3

should be modified by simply replacing y_i^* with $\log(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2$ in Algorithm 1. For the prior of $\boldsymbol{\beta}$, we assume that $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta)$. Let $\mathbf{y} = (y_1, \ldots y_n)'$ and $\mathbf{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_n')'$. Then, it can be shown that

$$\boldsymbol{\beta}|\mathbf{y}, \mathbf{h} \sim N(\widehat{\boldsymbol{\mu}}_\beta, \widehat{\mathbf{V}}_\beta^{-1}), \tag{6.2}$$

where $\widehat{\mathbf{V}}_\beta = \mathbf{V}_\beta^{-1} + \mathbf{X}' \operatorname{Diag}(e^{-h_1}, \ldots, e^{-h_n})\mathbf{X}$ and $\widehat{\boldsymbol{\mu}}_\beta = \widehat{\mathbf{V}}_\beta^{-1}\left(\mathbf{V}_\beta^{-1}\boldsymbol{\mu}_\beta + \mathbf{X}' \operatorname{Diag}(e^{-h_1}, \ldots, e^{-h_n})\mathbf{y}\right)$. Thus, our auxiliary mixture sampler in Section 3 is completed with this extra block of sampling from $p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{h})$.

In our second extension, we consider a first-order spatial autoregressive process for the dependent variable. More precisely, we consider

$$y_i = \rho \sum_{j=1}^{n} m_{ij} y_j + x_i'\beta + \nu_i, \quad \nu_i = e^{\frac{1}{2}h_i}\varepsilon_i, \quad \text{for} \quad i = 1, \ldots, n, \tag{6.3}$$

where m_{ij}'s are exogenous spatial weights, ν_i is the regression disturbance term, h_i is the log-volatility of ν_i, and ρ is the scalar spatial autoregressive parameter measuring the degree of spatial correlation in the dependent variable. In vector form, (6.3) can be written as

$$\mathbf{y} = \rho\mathbf{M}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \operatorname{Diag}(e^{\frac{1}{2}h_1}, \ldots, e^{\frac{1}{2}h_n})\boldsymbol{\varepsilon}, \tag{6.4}$$

where $\mathbf{M} = (m_{ij})$ is the $n \times n$ matrix of spatial weights that has zero diagonal elements. Let $\mathbf{R}(\rho) = (\mathbf{I}_n - \rho\mathbf{M})$. Since the spatial autoregressive model is an equilibrium model, we assume that the following reduced form exits,

$$\mathbf{y} = \mathbf{R}^{-1}(\rho)\mathbf{X}\boldsymbol{\beta} + \mathbf{R}^{-1}(\rho) \operatorname{Diag}(e^{\frac{1}{2}h_1}, \ldots, e^{\frac{1}{2}h_n})\boldsymbol{\varepsilon}. \tag{6.5}$$

Next, we consider the extension of our auxiliary mixture sampler in Algorithm 1 for this model. First, the auxiliary sampler should be modified by replacing y_i^* with $\log(y_i - \rho\sum_{j=1}^n m_{ij}y_j - \mathbf{x}_i'\boldsymbol{\beta})^2$ in sampling $\mathbf{s}$, $\mathbf{h}$, σ_u^2, μ_h and λ in Algorithm 1. Second, we assume the following prior distributions for $\boldsymbol{\beta}$ and ρ to determine the conditional posteriors of these parameters: $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta)$, and $\rho \sim \text{Uniform}(-1/\gamma, 1/\gamma)$, where γ is the spectral radius of $\mathbf{M}$. Then, we have

$$\begin{aligned} p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{h}, \rho) \propto & \exp\left(-\frac{1}{2}(\mathbf{R}(\rho)\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \operatorname{Diag}(e^{-h_1}, \ldots, e^{-h_n})(\mathbf{R}(\rho)\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \\ & \times \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)'\mathbf{V}_\beta^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)\right), \end{aligned} \tag{6.6}$$

which implies that

$$\boldsymbol{\beta}|\mathbf{y}, \mathbf{h}, \rho \sim N(\widehat{\boldsymbol{\mu}}_\beta, \widehat{\mathbf{V}}_\beta^{-1}), \tag{6.7}$$

where $\widehat{\mathbf{V}}_\beta = \mathbf{V}_\beta^{-1} + \mathbf{X}' \operatorname{Diag}(e^{-h_1}, \ldots, e^{-h_n})\mathbf{X}$ and $\widehat{\boldsymbol{\mu}}_\beta = \widehat{\mathbf{V}}_\beta^{-1}\left(\mathbf{V}_\beta^{-1}\boldsymbol{\mu}_\beta + \mathbf{X}' \operatorname{Diag}(e^{-h_1}, \ldots, e^{-h_n})\mathbf{R}(\rho)\mathbf{y}\right)$. Finally, the conditional posterior distribution of ρ is given by

$$p(\rho|\mathbf{y}, \mathbf{h}, \boldsymbol{\beta}) \propto |\mathbf{R}(\rho)| \times \exp\left(-\frac{1}{2}(\mathbf{R}(\rho)\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \operatorname{Diag}(e^{-h_1}, \ldots, e^{-h_n})(\mathbf{R}(\rho)\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \tag{6.8}$$

which is not in a standard form. The random walk Metropolis-Hastings algorithm discussed in Section 4 can be used to generate random draws from $p(\rho|\mathbf{y}, \mathbf{h}, \boldsymbol{\beta})$.

For the final extension, we consider a model that has spatial dependence in both the dependent variable and the disturbance terms. More specifically, we consider

$$y_i = \rho_1 \sum_{j=1}^{n} m_{1,ij} y_j + \mathbf{x}_i' \boldsymbol{\beta} + \nu_i, \quad \nu_i = \rho_2 \sum_{j=1}^{n} m_{2,ij} \nu_j + e^{\frac{1}{2}h_i}\varepsilon_i, \quad i = 1, \ldots, n, \tag{6.9}$$

where $m_{1,ij}$ and $m_{2,ij}$ are spatial weights, ν_i is the regression disturbance term and the scalar parameters ρ_1 and ρ_2 are spatial autoregressive parameters. In vector form, we have

$$\mathbf{y} = \rho_1 \mathbf{M}_1 \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} = \rho_2 \mathbf{M}_2 \boldsymbol{\nu} + \operatorname{Diag}(e^{\frac{1}{2}h_1}, \ldots, e^{\frac{1}{2}h_n})\boldsymbol{\varepsilon}, \tag{6.10}$$

where $\mathbf{M}_1 = (m_{1ij})$ and $\mathbf{M}_2 = (m_{2ij})$ are $n \times n$ spatial weights matrices that have zero diagonal elements, and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)'$ is the $n \times 1$ vector of disturbance terms. Let $\mathbf{R}_1(\rho_1) = (\mathbf{I}_n - \rho_1 \mathbf{M}_1)$ and $\mathbf{R}_2(\rho_2) = (\mathbf{I}_n - \rho_2 \mathbf{M}_2)$. Under the assumption that $\mathbf{R}_1(\rho_1)$ and $\mathbf{R}_2(\rho_2)$ are invertible, the reduced form of (6.9) is given by

$$\mathbf{y} = \mathbf{R}_1^{-1}(\rho_1)\mathbf{X}\boldsymbol{\beta} + \mathbf{R}_1^{-1}(\rho_1)\mathbf{R}_2^{-1}(\rho_1) \operatorname{Diag}(e^{\frac{1}{2}h_1}, \ldots, e^{\frac{1}{2}h_n})\boldsymbol{\varepsilon}. \tag{6.11}$$

We note that (6.11) is in the form of (6.5) and therefore the same estimation approach can be adopted for the estimation of this extended model.

7 A Monte Carlo Study

7.1 Design

In this section, we design a Monte Carlo study to assess sampling properties of the suggested Bayesian algorithms. We consider two different data generating processes (DGPs). We will refer to the first case as the SV specification, and to the latter as the SARSV specification. More specifically,

the DGPs respectively are

$$y_i = e^{\frac{1}{2}h_i}\,\varepsilon_i, \quad h_i - \mu_h = \lambda \sum_{j=1}^{n} w_{ij}(h_j - \mu_h) + u_i, \quad \text{for} \quad i = 1, \ldots, n, \tag{7.1}$$

$$y_i = \rho \sum_{j=1}^{n} m_{ij} y_j + \mathbf{x}_i^{'}\boldsymbol{\beta} + e^{\frac{1}{2}h_i}\,\varepsilon_i, \quad h_i - \mu_h = \lambda \sum_{j=1}^{n} w_{ij}(h_j - \mu_h) + u_i, \quad \text{for} \quad i = 1, \ldots, n, \tag{7.2}$$

where ε_i's and u_i's are generated independently from the standard normal distribution. For the weights matrices, we use the two cases that we consider in our empirical application section. We generate rook and queen contiguity-based weights matrices from the 1292 census tracts in the broader Chicago Metropolitan area. Both weights matrices are row normalized so that each row sums to unity. In the SARSV specification, we set $\mathbf{W} = \mathbf{M}$ and only include an intercept term as the explanatory variable, i.e. $x_i = 1$ for all i. Given our parameter estimates from our empirical application in Section 8, we use the following true parameter values: $\rho = 0.15$, $\beta = 0.05$, $\lambda = 0.9$, $\mu_h = -3$ and $\sigma_u^2 = 0.5$. The number of resamples for both experiments is set to 50.

For the prior distributions, we consider the following: $\sigma_u^2 \sim$ IG(2, 0.5), $\rho \sim$ Uniform$(-1, 1)$, $\lambda \sim$ Uniform$(-1, 1)$, $\mu_h \sim N(0, 10)$ and $\beta \sim N(0, 10)$. The length of the Markov chain is 12000 draws, and the first 2000 draws are discarded to dissipate the effect of the initial values. We consider the following algorithms: (i) Algorithm 1 with the random-walk MH sampler for ρ and λ (10_MH), (ii) Algorithm 2 with the random-walk MH sampler for ρ and λ (N_MH). To determine adequacy of the length of our samplers, we apply the methodology suggested by Raftery and Lewis (1992). Some exemplary trace plots of the Bayesian estimates are provided in Figures 3–5 to demonstrate the convergence of MCMC samplers. In these trace plots, the red solid line corresponds to the estimated posterior mean, 10_MH denotes Algorithm 1 and N_MH denotes Algorithm 2. For the sake of brevity, we only present the results for the Queen contiguity case for the SARSV model. These trace plots show that the estimated posterior means are close to the corresponding true parameter values. Following Jacquier et al. (2004), we report the average bias and the root mean square errors (RMSE) for ρ, β, λ, μ_h and σ_u^2, and the following two summary measures for h. Let h_{ji}^r be the ith log-volatility calculated in the jth repetition and rth-pass from the sampler. Then, for the stochastic volatility, we report the grand RMSE calculated by RMSE$= \left[\frac{1}{50\times R\times n}\sum_{i=1}^{n}\sum_{j=1}^{50}\sum_{r=1}^{R}(h_{ji}^r - \widehat{h}_{ji})^2\right]^{1/2}$. We also report the mean absolute errors calculated by MAE$= \frac{1}{n}\sum_{i=1}^{n}|\widehat{h}_i - h_i|$, where $\widehat{h}_i$ is the Bayesian estimate of ith log-volatility and h_i is the corresponding true value. For the model selection, we focus on the true model selection frequency over 50 repetitions. To this end, we calculate the rejection frequency for the null hypothesis $H_0 : \lambda = 0.9$ using the associated SDDR's based on Algorithm 3.

7.2 Simulation Results

The simulation results are presented in Table 2 for the SV model and Table 3 for the SARSV model. In these tables, 10_MH denotes the auxiliary mixture sampler in Algorithm 1 and N_MH denotes the naive sampler in Algorithm 2. In terms of scalar measures reported in Table 2, we observe that both algorithms perform similarly for λ and σ_u^2 in terms of bias and efficiency. The estimated posterior means for these parameters are close their true values in both algorithms. However, the naive sampler (Algorithm 2) imposes severe downward bias on μ_h, which is about 80%. This also results in a significant loss in efficiency as shown by the corresponding large RMSE. Similarly, for $\mathbf{h}$, the two summary measures indicate that the naive sampler performs worse than the auxiliary mixture sampler (Algorithm 1) as the auxiliary mixture sampler reports relatively much smaller MAE and RMSE. The simulation results are similar for both queen and rook weights matrices, indicating that the performances of algorithms are invariant to spatial weights matrix specifications.

The simulation results for λ, σ_u^2 and $\mathbf{h}$ in Tables 3 are similar to those obtained for the SV model. Again, we observe that the naive sampler imposes severe downward bias on μ_h and performs worse than the auxiliary mixture sampler for $\mathbf{h}$. The performances of both samplers are similar for β, λ and σ_u^2. In the case of ρ, while the auxiliary mixture sampler imposes almost no bias, the naive sampler imposes significant positive bias. These results can also be confirmed from the exemplary trace plots given in Figures 3–5. For brevity, we only present the results for the Queen contiguity case for the SARSV model. For example, in the case of naive sampler, the red solid lines in these figures for ρ and μ_h are not close to the corresponding true parameter values. In Figure 5 (b), we provide plots in which we compare the estimated log-volatilities across all repetitions ($\widehat{\mathbf{h}}$) with the true log-volatilities. The estimates based on the auxiliary mixture sampler are very close to the true ones, whereas the estimates from the naive sampler are substantially smaller than the corresponding true values.

Finally, we compare the performance of both algorithms in terms of true model selection. Table 4 presents the frequency of rejecting the true null hypothesis, $H_0 : \lambda = 0.9$. In the queen contiguity case, we see that both samplers perform similarly with a rejection rate about 4% for the SARSV model, but the mixture auxiliary sampler slightly over rejects the null hypothesis for the SV model. When we look at the results for the rook contiguity case, both samplers significantly over reject the null hypothesis for the SV model, but for the SARSV model while the naive sampler has the rejection rate around 4%, the the auxiliary mixture sampler has the rejection rate around 28%.

Table 2: Simulation results for SV model

		$\widehat{\lambda}$		$\widehat{\sigma}_u^2$		$\widehat{\mu}_h$		$\widehat{\mathbf{h}}$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	MAE	RMSE
Queen	10_MH	0.029	0.034	-0.041	0.093	0.072	0.313	0.480	1.172
	N_MH	0.023	0.033	-0.004	0.133	-2.437	2.458	2.516	7.841
Rook	10_MH	0.027	0.032	-0.026	0.088	0.069	0.305	0.477	1.253
	N_MH	0.021	0.031	0.022	0.138	-2.454	2.474	2.531	8.042

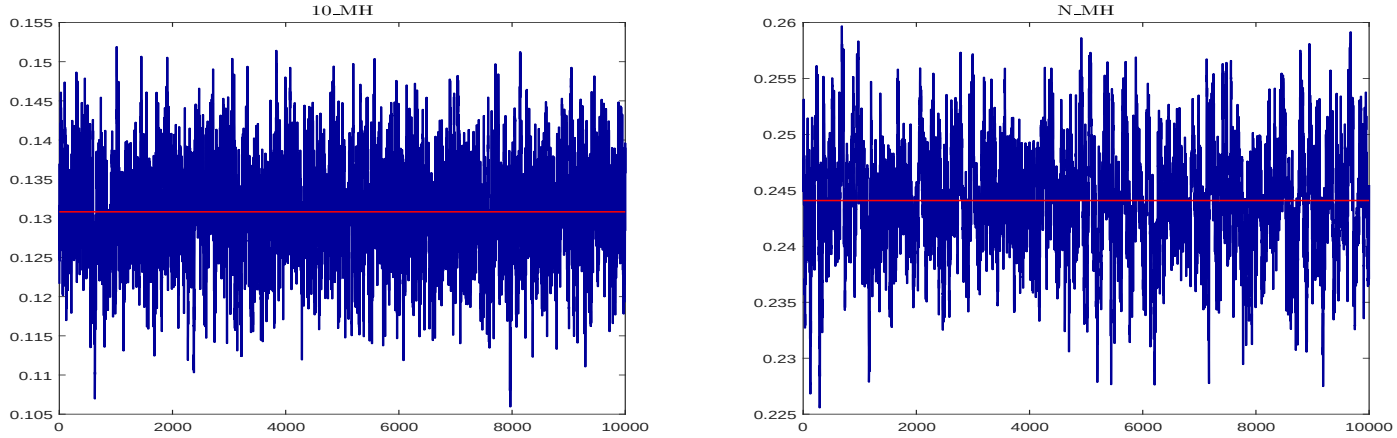
Notes: (i) 10_MH denotes Algorithm 1 and (ii) N_MH denotes Algorithm 3.

Table 3: Simulation results for SARSV model

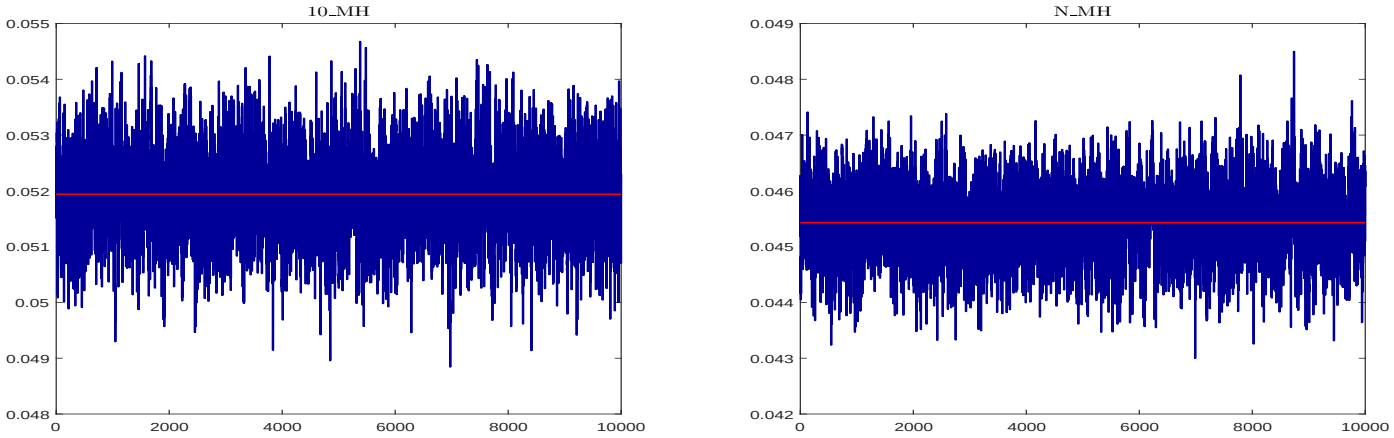
		$\widehat{\rho}$		$\widehat{\beta}$		$\widehat{\lambda}$		$\widehat{\sigma}_u^2$		$\widehat{\mu}_h$		$\widehat{\mathbf{h}}$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	MAE	RMSE
Queen	10_MH	-0.019	0.019	0.002	0.002	0.028	0.034	-0.032	0.091	0.059	0.309	0.476	1.174
	N_MH	0.094	0.094	-0.005	0.005	0.018	0.031	0.030	0.146	-2.466	2.487	2.541	7.999
Rook	10_MH	-0.003	0.003	0.001	0.001	0.029	0.034	-0.035	0.090	0.059	0.307	0.481	1.252
	N_MH	0.136	0.136	-0.007	0.007	0.020	0.031	0.014	0.135	-2.473	2.491	2.546	8.125

Notes: (i) 10_MH denotes Algorithm 1 and (ii) N_MH denotes Algorithm 3.

Figure 3: Trace plots for $\widehat{\rho}$ and $\widehat{\beta}$

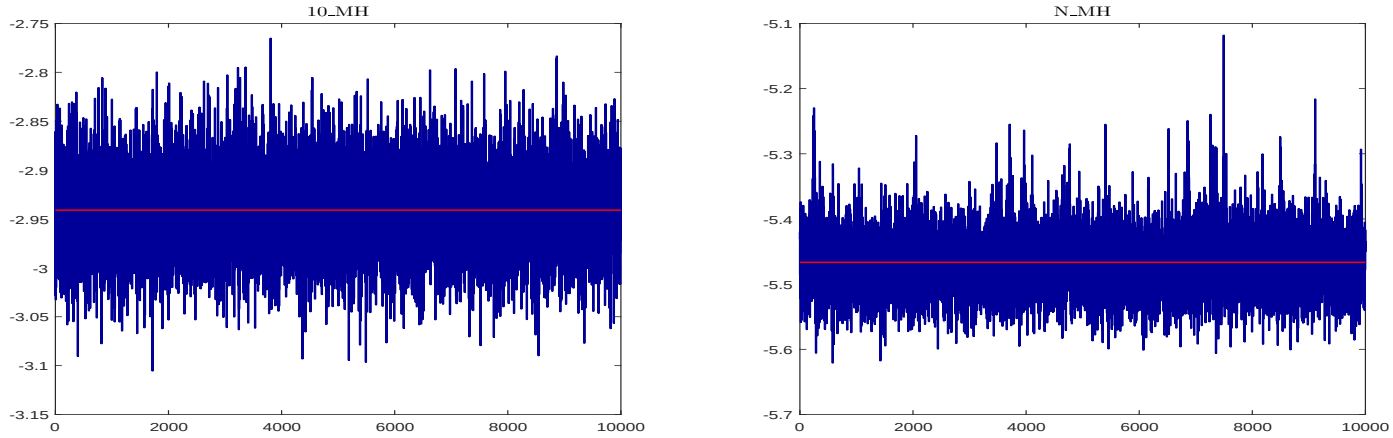


(a) Trace plots for $\widehat{\rho}$ when $\rho = 0.15$

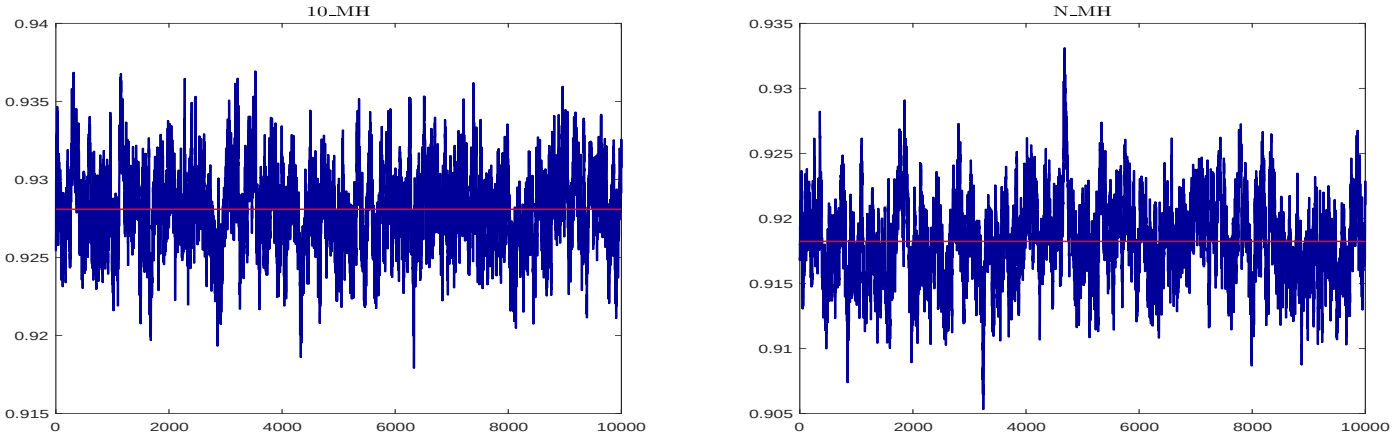


(b) Trace plots for $\widehat{\beta}$ when $\beta = 0.05$

Figure 4: Trace plots for $\widehat{\mu}_h$ and $\widehat{\lambda}$

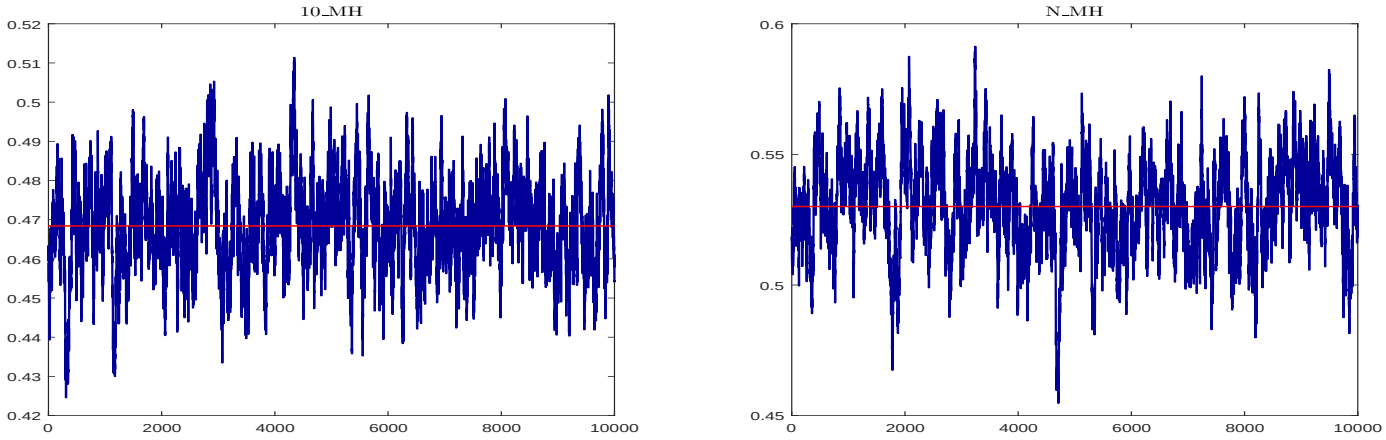


(a) Trace plots for $\widehat{\mu}_h$ when $\mu_h = -3$

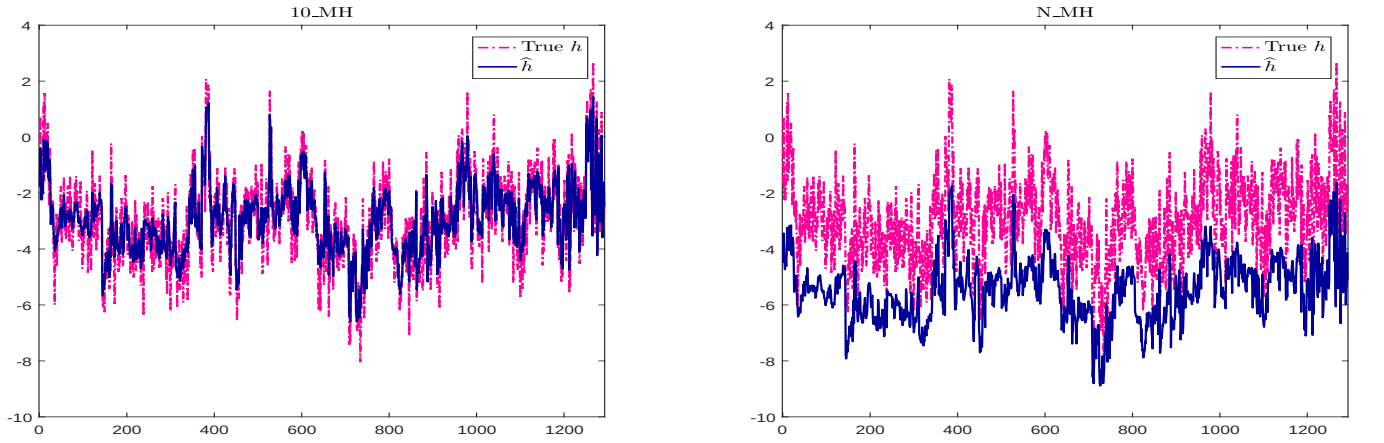


(b) Trace plots for $\widehat{\lambda}$ when $\lambda = 0.9$

Figure 5: Trace plots for $\widehat{\sigma}_u^2$ and estimated log-volatility



(a) Trace plots for $\widehat{\sigma}_u^2$ when $\sigma_u^2 = 0.5$



(b) $\widehat{\mathbf{h}}$ versus true $\mathbf{h}$

Table 4: Savage-Dicky Density Ratio: $\lambda = 0.9$

		SV		SARSV	
		10_MH	N_MH	10_MH	N_MH
Queen	frequency	0.080	0.040	0.040	0.040
	mean	0.301	0.296	0.245	0.285
	Sdev	0.387	0.540	0.262	0.705
Rook	frequency	0.360	0.240	0.280	0.040
	mean	1.184	2.863	2.515	0.453
	Sdev	1.820	9.792	7.546	1.227

Notes: (i) 10_MH denotes Algorithm 1 and (ii) N_MH denotes Algorithm 3.

8 An Empirical Application

We use the residential property sale prices in broader Chicago Metropolitan area for years 2014 and 2015. The data set is obtained from the Illinois Realtors (IAR) and contains annual median house prices for residential properties in 1292 census tracts over 38 townships. We calculate the annual returns from the first difference of log-annual median house prices. The summary statistics on returns are given in Table 5. The reported sample standard deviations indicate that West township has the largest variations in returns, followed by Hyde Park, Lake, Bloom, South and Calumet.

Before specifying the model, we use the Moran I test to check for existence of spatial correlation in the returns and squared returns. The Moran I statistic tests the null hypothesis of no spatial correlation against an unspecified form of spatial correlation. We use the queen and rook contiguity weights matrices for the calculation of the test statistic. The results are reported in Table 6. The results indicate that there is mild statistical evidence for the rejection of the null hypothesis of no spatial correlation in returns, and there is strong evidence for the spatial correlation in squared returns. However, the reported test statistics are substantially larger in the case of squared returns suggesting strong spatial dependence. The presence of spatial correlation in the return series indicates that the housing market may not be fully efficient in the sense that the return in a census tract can be predicted by the returns in the nearby census tracts. This result is not surprising since the frictions in a housing market such as transaction costs, carrying costs and tax considerations may limit arbitrage opportunities leading to pricing inefficiencies (Case and Shiller, 1989).

Given the non-zero overall mean in Table 5 and the Moran I test results, we consider the following version of (6.3) for the estimation

$$y_i = \mu + \rho \sum_{j=1}^{n} w_{ij} y_j + e^{\frac{1}{2}h_i}\,\varepsilon_i, \quad \text{where} \quad h_i = \mu_h + \lambda \sum_{j=1}^{n} w_{ij}(h_j - \mu_h) + u_i, \tag{8.1}$$

Table 5: Descriptive Statistics: Returns

	Nobv	Mean	Median	Sdv	Min	Max
Whole Sample	1292	0.096	0.068	0.336	-1.936	2.278
Townships						
Barrington	2	0.195	0.195	0.190	0.060	0.329
Berwyn	8	0.175	0.135	0.091	0.091	0.355
Bloom	32	-0.044	-0.009	0.398	-1.937	0.511
Bremen	13	0.072	0.058	0.210	-0.270	0.636
Calumet	11	0.040	0.014	0.287	-0.336	0.671
Cicero	10	0.041	0.080	0.224	-0.344	0.278
Elk Grove	7	0.100	0.126	0.060	-0.004	0.156
Evanston	14	0.044	0.033	0.237	-0.337	0.448
Hanover	16	0.148	0.106	0.121	-0.025	0.424
Hyde Park	140	0.116	0.094	0.496	-1.193	1.720
Jefferson	112	0.096	0.100	0.202	-1.075	0.509
Lake	125	0.108	0.090	0.407	-1.407	1.518
Lake View	100	0.039	0.018	0.203	-0.361	1.139
Lemont	2	-0.053	-0.053	0.045	-0.085	-0.022
Leyden	14	0.019	0.127	0.254	-0.633	0.257
Lyons	10	0.100	0.104	0.156	-0.170	0.431
Maine	22	0.110	0.074	0.138	-0.032	0.553
New Trier	13	0.009	-0.045	0.120	-0.140	0.248
Niles	27	0.059	0.058	0.090	-0.195	0.205
North	32	0.014	0.024	0.119	-0.292	0.339
Northfield	24	0.053	0.046	0.130	-0.218	0.274
Norwood Park	14	0.043	0.039	0.093	-0.122	0.245
Oak Park	6	0.039	0.031	0.253	-0.268	0.403
Orland	26	0.071	0.062	0.076	-0.046	0.255
Palatine	32	0.085	0.058	0.200	-0.431	0.909
Palos	10	0.084	0.093	0.109	-0.105	0.303
Proviso	49	0.169	0.103	0.205	-0.105	0.835
Rich	6	0.041	0.078	0.207	-0.208	0.254
River Forest	2	-0.112	-0.112	0.267	-0.301	0.076
Riverside	2	-0.032	-0.032	0.091	-0.096	0.033
Rogers Park	30	0.157	0.138	0.238	-0.321	0.797
Schaumburg	26	0.101	0.061	0.213	-0.248	0.926
South	49	0.104	0.033	0.328	-0.583	1.151
Stickney	24	0.096	0.083	0.144	-0.080	0.604
Thornton	32	0.030	0.088	0.178	-0.458	0.277
West	167	0.186	0.112	0.566	-1.754	2.278
Wheeling	34	0.069	0.045	0.162	-0.189	0.779
Worth	49	0.060	0.068	0.112	-0.237	0.298

for $i = 1, \ldots, 1292$, where μ is the constant mean. We do not consider the effect of location and housing attributes on the level of returns in (8.1). The literature on the decompositions of temporal changes in house prices shows that house prices are mainly driven by altered coefficients in hedonic regressions rather than by the changes in characteristics (McMillen, 2008; Nicodemo and Raya,

Table 6: Moran I test results

	Queen	Rook
Returns	0.037	0.032
	(0.009)	(0.034)
Squared Returns	0.188	0.184
	(0.000)	(0.000)

Notes: P-values are reported in parentheses.

2012; Thomschke, 2015). In particular, McMillen (2008) shows that the variables on location and housing attributes do not explain the change in the house price distributions by using a sample of sales prices of single-family homes in Chicago for the years 1995 and 2005.

In (8.1), we specify two weights matrices according to queen and rook contiguity. Both weights matrices are row normalized. We use the measure of fit called the deviance information criterion (DIC) suggested by Spiegelhalter et al. (2002) to determine which weights matrix is more compatible with the sample data. In the context of our model, the DIC is given by

$$\text{DIC} = -4\,\mathbb{E}\left[\log p\left(\mathbf{y}^*|\mathbf{s},\mathbf{h}\right)\big|\mathbf{y}^*\right] + 2\log p(\mathbf{y}^*|\widehat{\mathbf{s}},\widehat{\mathbf{h}}), \tag{8.2}$$

where $p\left(\mathbf{y}^*|\mathbf{s},\mathbf{h}\right)$ is the conditional likelihood given in (4.4), and $\widehat{\mathbf{s}}$ and $\widehat{\mathbf{h}}$ are the estimates of $\mathbf{s}$ and $\mathbf{h}$, respectively. The first term $\mathbb{E}\left[\log p\left(\mathbf{y}^*|\mathbf{s},\mathbf{h}\right)\big|\mathbf{y}^*\right]$ can be estimated by averaging the conditional likelihood $\log p\left(\mathbf{y}^*|\mathbf{s},\mathbf{h}\right)$ over the posterior draws of $\mathbf{s}$ and $\mathbf{h}$. Then, given a set of competing models each of which corresponds to a different weights matrix, the preferred model is the one with the minimum DIC value.[5]

Table 7: Analysis of Chicago Area House Sales Returns

		$\widehat{\rho}$	$\widehat{\mu}$	$\widehat{\mu}_h$	$\widehat{\lambda}$	$\widehat{\sigma_u^2}$	$\widehat{h}$
Queen	mean	0.145	0.052	-3.305	0.908	0.542	-3.313
	median	0.146	0.052	-3.307	0.910	0.535	-3.566
	sdev	0.044	0.005	0.247	0.021	0.108	0.790
Rook	mean	0.114	0.055	-3.269	0.916	0.433	-3.275
	median	0.114	0.055	-3.272	0.917	0.422	-3.514
	sdev	0.044	0.005	0.239	0.020	0.093	0.764
DIC (Queen)	4025.284						
DIC (Rook)	3955.915						

Notes: (i) $\widehat{h}$ is the mean of estimated log-volatilities, and (ii) DIC stands for the deviance information criterion.

[5]The DIC in (8.2) is called the conditional DIC , since it is formulated with the conditional likelihood $p\left(\mathbf{y}^*|\mathbf{s},\mathbf{h}\right)$ (Berg et al., 2004; Celeux et al., 2006). Note that the conditional DIC can favor overfitted models in model comparison exercises (Chan and Grant, 2016).

Table 8: Means of estimated conditional standard deviations by township

		Queen					Rook				
Township	Nobv	Mean	Median	Sdv	Min	Max	Mean	Median	Sdv	Min	Max
Barrington	2	0.166	0.166	0.018	0.153	0.178	0.163	0.163	0.014	0.153	0.173
Berwyn	8	0.139	0.137	0.050	0.087	0.239	0.139	0.129	0.059	0.087	0.267
Bloom	32	0.249	0.237	0.113	0.127	0.762	0.255	0.232	0.114	0.132	0.753
Bremen	13	0.153	0.141	0.049	0.102	0.289	0.155	0.141	0.049	0.102	0.283
Calumet	11	0.285	0.274	0.087	0.172	0.439	0.291	0.279	0.087	0.158	0.466
Cicero	10	0.309	0.307	0.084	0.208	0.464	0.324	0.326	0.099	0.205	0.481
Elk Grove	7	0.086	0.085	0.008	0.076	0.095	0.089	0.090	0.007	0.079	0.097
Evanston	14	0.219	0.215	0.056	0.152	0.312	0.213	0.204	0.051	0.151	0.295
Hanover	16	0.148	0.144	0.027	0.108	0.198	0.153	0.154	0.028	0.112	0.203
Hyde Park	140	0.475	0.464	0.151	0.148	1.026	0.475	0.460	0.148	0.149	0.989
Jefferson	112	0.184	0.180	0.076	0.080	0.579	0.188	0.180	0.078	0.079	0.602
Lake	125	0.354	0.339	0.176	0.054	0.800	0.363	0.357	0.174	0.055	0.806
Lake View	100	0.164	0.156	0.070	0.062	0.514	0.167	0.154	0.074	0.064	0.519
Lemont	2	0.117	0.117	0.006	0.112	0.121	0.113	0.113	0.006	0.108	0.117
Leyden	14	0.177	0.168	0.071	0.106	0.360	0.177	0.164	0.065	0.113	0.338
Lyons	10	0.149	0.143	0.028	0.113	0.208	0.151	0.144	0.031	0.111	0.213
Maine	22	0.122	0.111	0.046	0.074	0.267	0.123	0.107	0.044	0.079	0.258
New Trier	13	0.149	0.146	0.015	0.124	0.176	0.150	0.148	0.015	0.129	0.183
Niles	27	0.109	0.101	0.022	0.084	0.168	0.107	0.101	0.023	0.085	0.171
North	32	0.135	0.130	0.025	0.096	0.210	0.137	0.136	0.023	0.101	0.205
Northfield	24	0.134	0.135	0.041	0.068	0.225	0.137	0.133	0.041	0.071	0.232
Norwood Park	14	0.116	0.114	0.026	0.078	0.161	0.119	0.114	0.030	0.079	0.182
Oak Park	6	0.232	0.219	0.049	0.173	0.311	0.243	0.244	0.042	0.183	0.305
Orland	26	0.087	0.090	0.023	0.048	0.141	0.087	0.090	0.022	0.048	0.131
Palatine	32	0.136	0.110	0.066	0.073	0.367	0.139	0.122	0.062	0.075	0.352
Palos	10	0.129	0.126	0.015	0.106	0.159	0.130	0.127	0.017	0.101	0.159
Proviso	49	0.191	0.161	0.098	0.063	0.456	0.191	0.171	0.096	0.068	0.444
Rich	6	0.227	0.232	0.024	0.190	0.257	0.224	0.229	0.024	0.185	0.253
River Forest	2	0.180	0.180	0.061	0.137	0.223	0.190	0.190	0.045	0.158	0.222
Riverside	2	0.135	0.135	0.003	0.133	0.137	0.134	0.134	0.004	0.131	0.137
Rogers Park	30	0.237	0.237	0.064	0.106	0.378	0.239	0.232	0.067	0.108	0.380
Schaumburg	26	0.164	0.163	0.058	0.089	0.369	0.170	0.168	0.061	0.089	0.361
South	49	0.313	0.275	0.162	0.097	0.725	0.314	0.277	0.155	0.099	0.675
Stickney	24	0.127	0.119	0.043	0.056	0.250	0.126	0.126	0.044	0.059	0.241
Thornton	32	0.195	0.191	0.039	0.123	0.305	0.197	0.190	0.037	0.139	0.304
West	167	0.442	0.353	0.286	0.096	1.442	0.444	0.364	0.282	0.099	1.424
Wheeling	34	0.117	0.114	0.042	0.068	0.305	0.117	0.115	0.041	0.070	0.294
Worth	49	0.126	0.121	0.031	0.088	0.229	0.127	0.118	0.030	0.090	0.229

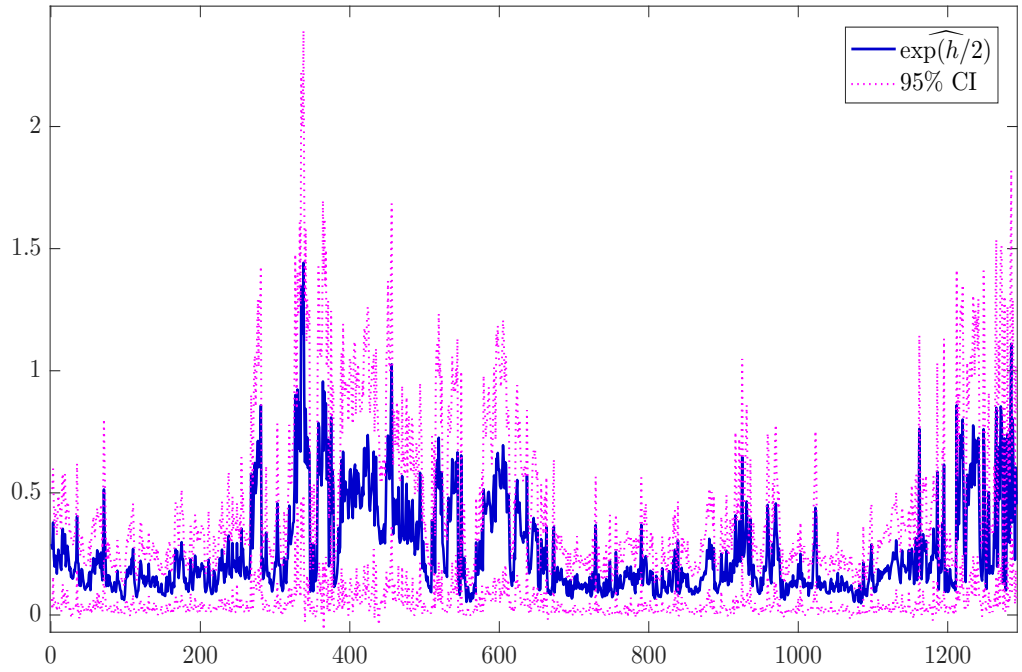
Notes: The mean of estimated conditional variance in a township is calculated by taking the average of estimated conditional variances in its census tracts.

The estimation results are reported in Table 7. The results are similar for both weights matrices and the DIC statistic indicates that the model based on the rook contiguity is more compatible with the sample data. Therefore, we focused on the rook-based results. The estimates of spatial autoregressive parameters ρ and λ are significant, and given by 0.114 and 0.916, respectively. These estimates indicate that though there is weak spatial correlation in returns, log-volatilities exhibit strong spatial correlation. The mean of estimated log-volatilities is -3.275, implying an average conditional standard deviation estimate of 0.1945. In Figure 6 shows the line plots of conditional variance estimates. These plots indicate that observations around 400 have relative larger conditional variance estimates. Table 8 gives the spatial distribution of average conditional variance estimates over townships. The five townships in order of large conditional variance estimates are West, Hyde Park, Lake, South and Calumet. Note that these townships also have relatively large variation in returns as shown in Table 5. The estimated conditional variances over census tracts are displayed in maps given in Figure 7. The lowest conditional variance estimates are distributed over the suburbs of the city of Chicago. On the other hand, the estimated conditional variances are relatively higher in the corridor extending from the west side of city to the south side of city. As shown in the maps, the spatial clustering in the west side of city has the highest estimates.

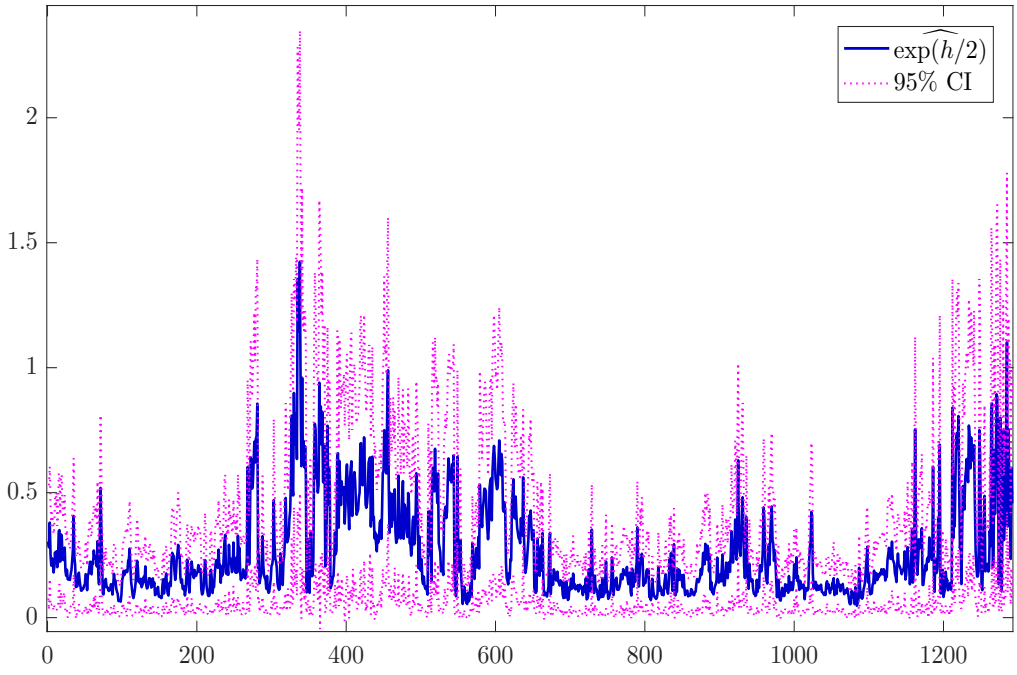
9 Conclusion

In this paper, we suggested a spatial stochastic volatility model that allows for a first-order spatial autoregressive process for the latent log-volatility. Our parsimonious model allowed us to make a clear distinction between spatial *dependence* and spatial *correlation*, which are often loosely used in the literature. We also proposed some extensions that allow for spatial correlation in the outcome equation by introducing the spatial lag of the dependent variable. For the estimation, we transformed the model so that it took the form of a linear Gaussian state-space model, where the log-volatility equation can be considered as the state equation. We devised Bayesian MCMC algorithms coupled with the data augmentation method for the estimation of parameters and the latent log-volatility terms. Our simulation results indicated that the Bayesian estimator using the Gaussian mixture approximation has good finite sample properties. In an empirical application using the price returns in the residential properties in the city of Chicago and its surrounding suburbs, we showed that although there is weak positive spatial correlation in the returns, but a strong positive spatial correlation in log-volatility terms. Our results on the estimated conditional variances indicated that the lowest estimates are distributed over the suburbs of the city of Chicago, while relatively larger estimates are distributed over a corridor extending from the west side of city to the south side.

Our analysis suggests a number of directions for future research. First, we have already considered some extensions of our parsimonious model in Section 6. In future research, Bayesian model comparison criteria based on either the marginal likelihood or integrated likelihood can be developed for model selection exercises. Second, panel data extensions of our model can be developed

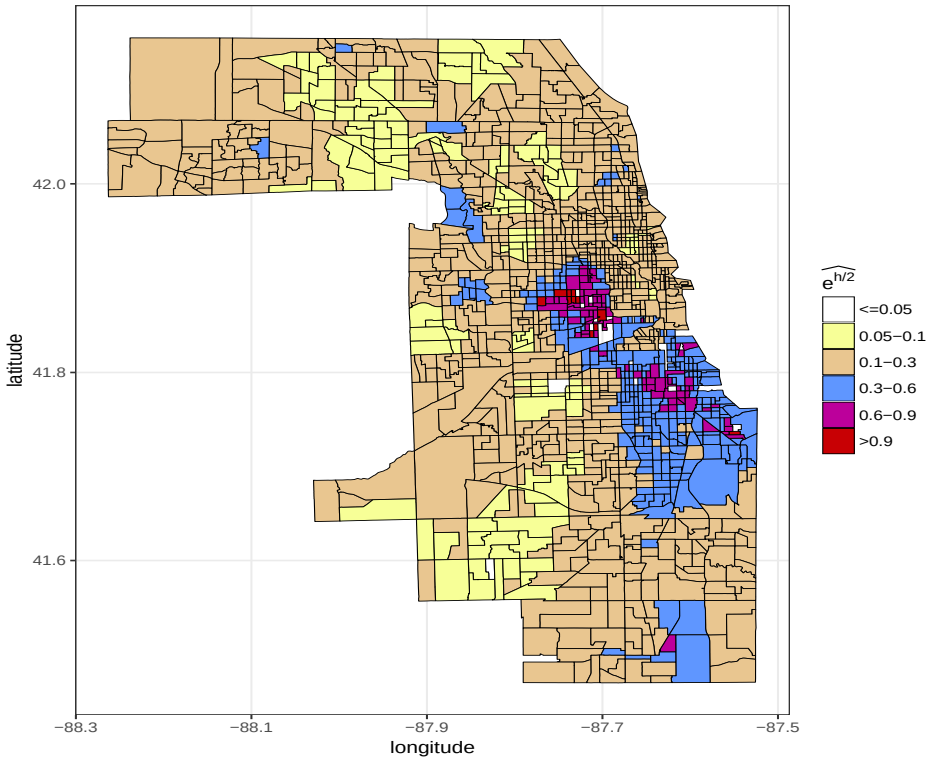


(c) Queen weights matrix

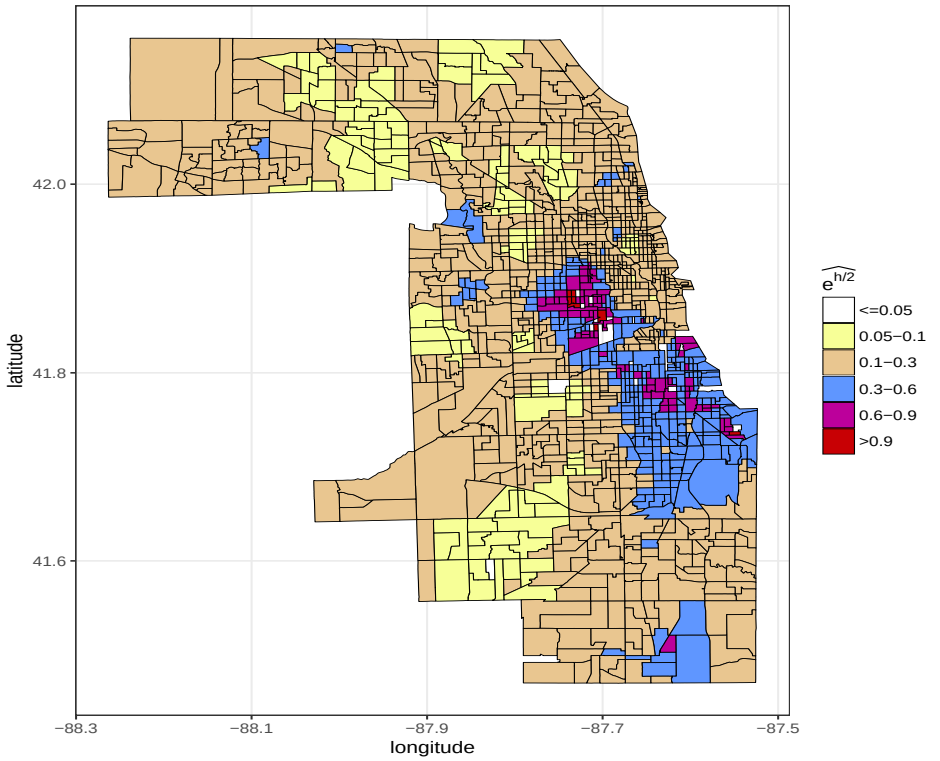


(d) Rook weights matrix

Figure 6: The line plots of conditional variance estimates



(a) Estimates based on the queen weights matrix



(b) Estimates based on the rook weights matrix

Figure 7: The conditional variance estimates

to allow for a rich structure in the outcome and log-volatility equations. Similarly, another fruitful avenue for future research would be the extension of our model to a multivariate model in which a correlation structure between log-volatility terms of different processes is allowed analogous to the multivariate generalized ARCH (GARCH) type models suggested by Bollerslev (1990) and Harvey et al. (1994). Last but not least, another important area for future research is to consider spatial panel data models in which weak and strong spatial correlations (Chudik et al., 2011) are allowed either for the outcome or the log-volatility equation.

References

Anselin, Luc (1988). *Spatial econometrics: Methods and Models*. New York: Springer.

— (2007). "Spatial Econometrics". In: *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*. Ed. by Kerry Patterson and Terence C. Mills. Palgrave Macmillan.

Bailey, Natalia, Sean Holly, and M. Hashem Pesaran (2016). "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence". In: *Journal of Applied Econometrics* 31.1, pp. 249–280.

Baltagi, Badi H., Bernard Fingleton, and Alain Pirotte (2014). "Estimating and Forecasting with a Dynamic Spatial Panel Data Model*". In: *Oxford Bulletin of Economics and Statistics* 76.1, pp. 112–138.

Beenstock, Michael and Daniel Felsenstein (2007). "Spatial Vector Autoregressions". In: *Spatial Economic Analysis* 2.2, pp. 167–196.

Berg, Andreas, Renate Meyer, and Jun Yu (2004). "Deviance Information Criterion for Comparing Stochastic Volatility Models". In: *Journal of Business & Economic Statistics* 22.1, pp. 107–120.

Bollerslev, Tim (1990). "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model". In: *The Review of Economics and Statistics* 72.3, pp. 498–505.

Brady, Ryan R. (2011). "Measuring the diffusion of housing prices across space and over time". In: *Journal of Applied Econometrics* 26.2, pp. 213–231.

Case, Karl E. and Robert J. Shiller (1989). "The Efficiency of the Market for Single-Family Homes". In: *The American Economic Review* 79.1, pp. 125–137.

Celeux, G. et al. (Dec. 2006). "Deviance information criteria for missing data models". In: *Bayesian Analysis* 1.4, pp. 651–673.

Chan, Joshua C. C. and Angelia L. Grant (2016). "On the Observed-Data Deviance Information Criterion for Volatility Modeling". In: *Journal of Financial Econometrics* 14.4, pp. 772–802.

Chan, Joshua C.C. and Ivan Jeliazkov (2009). "Efficient simulation and integrated likelihood estimation in state space models". In: *International Journal of Mathematical Modelling and Numerical Optimisation* 1.1-2, pp. 101–120.

Chib, Siddhartha and Edward Greenberg (1994). "Bayes inference in regression models with ARMA (p, q) errors". In: *Journal of Econometrics* 64.1, pp. 183 –206.

— (1995). "Understanding the Metropolis-Hastings Algorithm". In: *The American Statistician* 49.4, pp. 327–335.

Chudik, Alexander, M. Hashem Pesaran, and Elisa Tosetti (2011). "Weak and strong cross-section dependence and estimation of large panels". In: *The Econometrics Journal* 14.1.

Cliff, Andrew and Keith Ord (1972). "Testing for Spatial Autocorrelation Among Regression Residuals". In: *Geographical Analysis* 4.3, pp. 267–284.

Dickey, James M. (1971). "The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters". In: *The Annals of Mathematical Statistics* 42.1, pp. 204–223.

Doğan, Osman and Süleyman Taşpınar (2018). "Bayesian Inference in Spatial Sample Selection Models". In: *Oxford Bulletin of Economics and Statistics* 80.1, pp. 90–121.

Elhorst, J. Paul (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer Briefs in Regional Science. New York: Springer Berlin Heidelberg.

Fridman, Moshe and Lawrence Harris (1998). "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models". In: *Journal of Business & Economic Statistics* 16.3, pp. 284–291.

Gallin, Joshua (2006). "The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets". In: *Real Estate Economics* 34.3, pp. 417–438.

Gelfand, Alan E. and Adrian F. M. Smith (1990). "Sampling-Based Approaches to Calculating Marginal Densities". In: *Journal of the American Statistical Association* 85.410, pp. 398–409.

Han, Xiaoyi and Lung-Fei Lee (2016). "Bayesian Analysis of Spatial Panel Autoregressive Models With Time-Varying Endogenous Spatial Weight Matrices, Common Factors, and Random Coefficients". In: *Journal of Business & Economic Statistics* 34.4, pp. 642–660.

Harvey, Andrew, Esther Ruiz, and Neil Shephard (1994). "Multivariate Stochastic Variance Models". In: *The Review of Economic Studies* 61.2, pp. 247–264.

Holly, Sean, M. Hashem Pesaran, and Takashi Yamagata (2010). "A spatio-temporal model of house prices in the USA". In: *Journal of Econometrics* 158.1, pp. 160 –173.

— (2011). "The spatial and temporal diffusion of house prices in the UK". In: *Journal of Urban Economics* 69.1, pp. 2 –23.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi (1994). "Bayesian Analysis of Stochastic Volatility Models". In: *Journal of Business & Economic Statistics* 12.4, pp. 371–389.

— (2004). "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors". In: *Journal of Econometrics* 122.1, pp. 185 –212.

Kass, Robert E. and Adrian E. Raftery (1995). "Bayes Factors". In: *Journal of the American Statistical Association* 90.430, pp. 773–795.

Kelejian, Harry H. and Ingmar R. Prucha (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances". In: *Journal of Econometrics* 157, pp. 53–67.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models". In: *The Review of Economic Studies* 65.3, pp. 361–393.

Kuethe, Todd H. and Valerien O. Pede (2011). "Regional Housing Price Cycles: A Spatio-temporal Analysis Using US State-level Data". In: *Regional Studies* 45.5, pp. 563–574.

Lee, Lung-fei (2004). "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models". In: *Econometrica* 72.6, pp. 1899–1925.

— (2007). "GMM and 2SLS estimation of mixed regressive, spatial autoregressive models". In: *Journal of Econometrics* 137.2, pp. 489–514.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin (2010). "Specification and estimation of social interaction models with network structures". In: *The Econometrics Journal* 13.2, pp. 145–176.

LeSage, James and Robert K. Pace (2009). *Introduction to Spatial Econometrics (Statistics: A Series of Textbooks and Monographs*. London: Chapman and Hall/CRC.

McMillen, Daniel P. (2008). "Changes in the distribution of house prices over time: Structural characteristics, neighborhood, or coefficients?" In: *Journal of Urban Economics* 64.3, pp. 573 –589.

Meen, Geoffrey (1996). "Spatial aggregation, spatial dependence and predictability in the UK housing market". In: *Housing Studies* 11.3, pp. 345–372.

— (1999). "Regional House Prices and the Ripple Effect: A New Interpretation". In: *Housing Studies* 14.6, pp. 733–753.

Nicodemo, Catia and Josep Maria Raya (2012). "Change in the distribution of house prices across Spanish cities". In: *Regional Science and Urban Economics* 42.4, pp. 739 –748.

Omori, Yasuhiro et al. (2007). "Stochastic volatility with leverage: Fast and efficient likelihood inference". In: *Journal of Econometrics* 140.2, pp. 425 –449.

Ord, Keith (1975). "Estimation Methods for Models of Spatial Interaction". In: *Journal of the American Statistical Association* 70.349, pp. 120–126.

Pesaran, M. Hashem (2015). "Testing Weak Cross-Sectional Dependence in Large Panels". In: *Econometric Reviews* 34.6-10, pp. 1089–1117.

Raftery, Adrian E. and Steven Lewis (1992). "How Many Iterations in the Gibbs Sampler?" In: *In Bayesian Statistics 4*. Ed. by J. M. Bernardo et al. Oxford University Press, pp. 763–773.

Ritter, Christian and Martin A. Tanner (1992). "Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler". In: *Journal of the American Statistical Association* 87.419, pp. 861–868.

Robinson, P. M. (2009). "Large-sample inference on spatial dependence". In: *Econometrics Journal* 12.

Ruiz, Esther (1994). "Quasi-maximum likelihood estimation of stochastic volatility models". In: *Journal of Econometrics* 63.1, pp. 289 –306.

Sandmann, Gleb and Siem Jan Koopman (1998). "Estimation of stochastic volatility models via Monte Carlo maximum likelihood". In: *Journal of Econometrics* 87.2, pp. 271 –301.

Shephard, Neil (1994). "Partial non-Gaussian state space". In: *Biometrika* 81.1, pp. 115–131.

Shi, Wei and Lung fei Lee (2017). "Spatial dynamic panel data models with interactive fixed effects". In: *Journal of Econometrics* 197.2, pp. 323 –347.

Spiegelhalter, David J. et al. (2002). "Bayesian measures of model complexity and fit". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.4, pp. 583–639.

Thomschke, Lorenz (2015). "Changes in the distribution of rental prices in Berlin". In: *Regional Science and Urban Economics* 51.Supplement C, pp. 88 –100.

Verdinelli, Isabella and Larry Wasserman (1995). "Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio". In: *Journal of the American Statistical Association* 90.430, pp. 614–618.