Analyzing Spatial Panel Data of Cigarette Demand:
A Bayesian Hierarchical Modeling Approach

Yanbing Zheng[1], Jun Zhu[2] and Dong Li[3]
[1] *University of Kentucky,* [2] *University of Wisconsin, Madison*
and [3] *Kansas State University*

Abstract: Analysis of spatial panel data is of great importance and inter-est in spatial econometrics. Here we consider cigarette demand in a spatial panel of 46 states of the US over a 30-year period. We construct a de-mand equation to examine the elasticity of per pack cigarette price and per capita disposable income. The existing spatial panel models account for both spatial autocorrelation and state-wise heterogeneity, but fail to account for temporal autocorrelation. Thus we propose new spatial panel models and adopt a fully Bayesian approach for model parameter inference and predic-tion of cigarette demand at future time points using MCMC. We conclude that the spatial panel model that accounts for state-wise heterogeneity, spa-tial dependence, and temporal dependence clearly outperforms the existing models. Analysis based on the new model suggests a negative cigarette price elasticity but a positive income elasticity.

Key words: Autoregressive model, demand equation, longitudinal data, Markov chain Monte Carlo, spatial econometrics, spatial-temporal process.

1. Introduction

In econometric terms, panel data are observations aggregated on a cross-section over multiple time periods. In the case of cross sections of spatial regions such as counties in a state, states in a country, or countries around the world, panel data are referred to as spatial panels (Anselin, 1988; Baltagi, 2005). Here we examine a demand equation that relates cigarette consumption to cigarette price and per capita disposable income. The federal government of the US has made serious effort through major policy interventions to reduce the consumption of cigarettes since the 1960's. However some of the major policy interventions are not without controversy, an example being the Congressional ban of broadcast advertising of cigarettes in 1971 after the application of the Fairness Doctrine Act to cigarette advertising in 1967. The banning of pro-smoking messages from

television and radio have eliminated anti-smoking messages, as it is no longer a requirement for these stations to adhere to the Fairness Doctrine Act. If anti-smoking messages are more effective than pro-smoking messages, then the net effect of the 1971 advertising ban may actually be an increase in the cigarette consumption. To assess this net effect and the effectiveness of other major policy interventions, it is important to properly construct and estimate the cigarette demand equation (Hamilton, 1972; Baltagi and Levin, 1986).
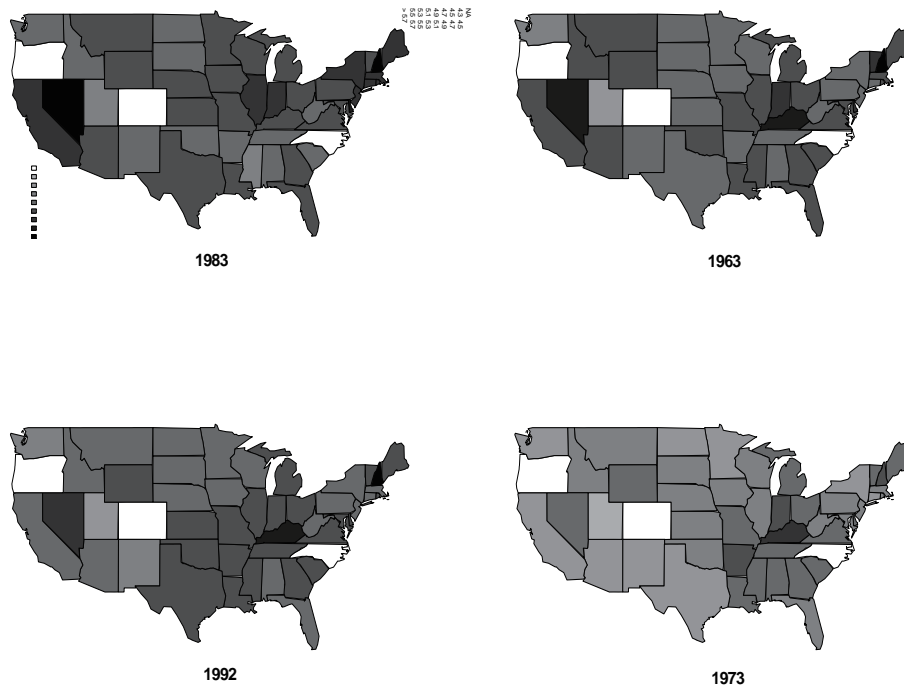

Figure 1: Gray-scaled maps of cigarette sales in a few selected years 1963, 1973, 1983 and 1992.

The data set consists of a panel of 46 states of the United States from 1963 to 1992. The response variable here is the real per capita sales of cigarettes by persons of smoking age (14 years and older) measured in packs per person on the log scale (henceforth "cigarette sales"). See Figure 1 for maps of cigarette sales in a few selected years. Two explanatory variables are considered for constructing the demand equation. The first explanatory variable is the average retail price of a pack of cigarettes measured in real terms on the log scale (henceforth "cigarette price"). The second explanatory variable is the real per capita disposable income of each state on the log scale (henceforth "income"). Following the convention in econometrics, all three variables, cigarette sales, cigarette price, and income, are on the log scale, so that the coefficients in the demand equation would represent elasticities.

While cigarette taxation is meant to deter cigarette consumption, it is still difficult to quantify the price elasticity of demand for cigarettes. One challenge is to account for spatial dependence between neighboring states, which may be attributed to economic phenomena such as bootlegging when consumers in a state of higher cigarette tax attempt to buy cigarettes in a bordering state of lower cigarette tax. Other challenges are to account for differences among different states or among different time points, which we will refer to as heterogeneity across states or heterogeneity over time. The effect of income on cigarette consumption can be ambiguous. On one hand, if cigarette is a normal good, higher income would lead to more consumption. On the other hand, more education, usually positively associated with higher income, would lead to less cigarette consumption. See Baltagi and Llevin (1986) and Baltagi and Li (2004) for more details.

Baltagi and levin (1986) were the first to consider this spatial panel from 1963 to 1980 and constructed a dynamic demand equation for cigarettes to address several major policy issues. Their data analysis yielded a significant negative effect of cigarette price on cigarette consumption with a price elasticity of -0.2, while there was no effect of income on cigarette consumption with an insignificant income elasticity. Furthermore, they modeled the bootlegging effect directly by incorporating the lowest cigarette price in neighboring states as an additional explanatory variable and found the effect of bootlegging to be statistically significant. In light of these data analysis results, they concluded that cigarette taxation was an effective tool for generating revenues despite the spillover effects to neighboring states where bootlegging was significant.

More recently, Baltagi and Li (2004) considered the same spatial panel with the time period updated to range from 1963 to 1992. They constructed a simple demand equation for cigarettes to examine the elasticity of cigarette price and that of income. They also modeled the bootlegging effect but as a part of spatial dependence, which may be thought of as an improvement of the explanatory variable approach taken in Baltagi and Levin (1986). In addition, depending on whether heterogeneity across states or over time was accounted for explicitly, the resulting models were spatially and temporally homogeneous, or temporally heterogeneous, or spatially heterogeneous. Based on the performance of prediction, they concluded that it was important to take into account spatial dependence and spatial heterogeneity.

There are, however, some unresolved issues in Baltagi and Li (2004). First, the models assume independence over time while allowing for dependence across space. It would be interesting and, as it turns out, necessary to account for both spatial dependence and temporal dependence for the cigarette demand data. One of the main reasons that temporal dependence is not accounted for in these spatial

panel models is due to computational difficulties involved in optimization over a high dimensional parameter space for obtaining the maximum likelihood estimates (MLE). Chen and Conley (2001) considered semiparametric estimation of spatial panel data models. Kapoor *et al.* (2006) proposed a method of moments estimator for a spatial panel data model. Driscoll and Kraay (1998) considered estimation of such models with unspecified spatial correlation. However, none of these models accounted for spatial dependence and temporal dependence simultaneously. See Elhorst (2003) and Anselin (2006) for a survey of spatial panel data models.

With the advances in computing technology and in Markov chain Monte Carlo (MCMC) algorithms, there is now ample opportunity to resolve these computational difficulties in spatial panel data analysis and consider more complex and thus more realistic spatial panel models. Here we consider generalizing the spatial panel models in Baltagi and Li (2004) to account for not only spatial dependence but also temporal dependence. In statistical terms, spatial panel data may be viewed as repeated measures of spatially aggregated data. Thus the models proposed here can be viewed as spatial-temporal statistical models for spatial panel data. For statistical inference, we use Bayesian hierarchical models as an alternative to maximum likelihood. We develop MCMC algorithms for obtaining the posterior distributions of model parameters, as well as posterior predictive distributions of cigarette demand at future time points.

Our contribution is that, for the cigarette demand data, our general spatial-temporal models provide a significant improvement over the existing spatial panel models in terms of model fitting as well as predictive power. We also show that the analysis results are comparable between MLE and Bayesian inference in the existing spatial panel models, while our Bayesian inference is still computationally feasible for the more complex spatial-temporal models. Furthermore, our approach provides direct statistical inference of spatial and temporal dependence in a panel data model, which, to our knowledge, has not been accomplished before. While Baltagi *et al.* (2006) considered the Lagrangian-Multiplier tests for similar models, these tests rely on the restricted model only and thus can not infer the unrestricted spatial-temporal model. In contrast, we provide the posterior distribution of spatial and temporal dependence in the unrestricted model. Finally, although spatial-temporal statistics have been applied to many disciplines such as ecology (Zhu *et al.* (2005, 2007) and epidemiology Waller *et al.* (1997), we believe that our spatial-temporal models and the Bayesian inference are novel for the analysis of spatial panel data in general. With greater availability of and a growing interest in spatial panel data, the work presented here would advance the capability of analyzing spatial panel data in practice and thus impact the field of spatial econometrics.

The remainder of the paper is organized as follows. We review the existing spatial panel models and propose new Bayesian inference in Section 2. Then in Section 3, we develop new spatial-temporal models for spatial panel data and again propose Bayesian inference. Analysis results of the cigarette demand data using both the existing and the new models are shown in Section 4. Model comparisons based on an information criterion and prediction performance are given in Section 5, followed by a brief conclusion in

2. Bayesian Inference for the Existing Spatial Panel Models

Here we review the spatial panel models considered in Baltagi and Li (2004) and propose a Bayesian hierarchical model for statistical inference. Following the notation in Baltagi and Li (2004), we let y_{it}, x_{it1}, x_{it2} denote the variables of cigarette sales, cigarette price, and income, respectively, in the ith state and the tth year, where $i = 1, \cdots, N$ indexes the $N = 46$ states and $t = 1, \ldots, T$ indexes the years starting from 1963. We will use the first $T = 25$ years (1963–1987) for model building and set aside the last 5 years (1988–1992) for prediction and model comparisons. The existing spatial panel model for cigarette demand is expressed as,

$$y_{it} = x'_{it}\beta + \epsilon_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T, \tag{2.1}$$

where $x_{it} = (x_{it1}, x_{it2})'$ denotes the two explanatory variables, $\beta = (\beta_1, \beta_2)'$ denotes the corresponding regression coefficients, and ϵ_{it} is such that

$$\epsilon_t = \mu_t + \phi_t, \tag{2.2}$$

where, at time t and in all states, $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$, $\mu_t = (\mu_{1t}, \ldots, \mu_{Nt})'$ denotes the vector of state effects, and $\phi_t = (\phi_{1t}, \ldots, \phi_{Nt})'$ denotes the vector of disturbance which are assumed to be independent of μ_t. Further, ϕ_t follows a spatial dependence model,

$$\phi_t = \lambda W \phi_t + \nu_t, \tag{2.3}$$

where W denotes an $N \times N$ matrix of known normalized spatial weights based on a pre-specified neighborhood structure, λ denotes a spatial autocorrelation coefficient where λ is between the inverse of the smallest eigenvalue of W and 1, and $\nu_t = (\nu_{1t}, \ldots, \nu_{Nt})'$ denotes white noise that are iid $N(0, \sigma_\nu^2)$ and are independent of ϕ_t and μ_t. The spatial dependence model in (2.3) is also known as the simultaneous autoregressive (SAR) model (Cressie, 1993). In Sections 2.1-2.4, we will consider different specifications of μ_t that give rise to four types of spatial panel models, which we will refer to as homogeneous model, heterogeneous

model, fixed-effects model, and random-effects model, similar to Baltagi and Li (2004).

Now let $y_t = (y_{1t}, \ldots, y_{Nt})'$ denote the vector of the response variables and let $X_t = [x_{1t}, \ldots, x_{Nt}]'$ denote the matrix of the explanatory variables in all states at time t. Further let $y = (y_1', \ldots, y_T')'$ and $X = [X_1', \ldots, X_T']'$ denote the response and explanatory variables in all states and at all times. Then we have the following distribution for the data at time t,

$$y_t | \beta, \mu_t, \sigma_\nu^2, \lambda \quad \sim \quad \text{normal}\left(X_t\beta + \mu_t, \sigma_\nu^2(I_N - \lambda W)^{-1}(I_N - \lambda W')^{-1}\right). \quad (2.4)$$

Under the assumption of temporal independence, the distribution for the data at all time points is,

$$y | \beta, \mu, \sigma_\nu^2, \lambda \quad \sim \quad \text{normal}\left(X\beta + \mu, \sigma_\nu^2 I_T \otimes (I_N - \lambda W)^{-1}(I_N - \lambda W')^{-1}\right) \quad (2.5)$$

where $\mu = (\mu_1', \ldots, \mu_T')'$. In Baltagi and Li (2004), statistical inference of the spatial panel models are via maximum likelihood and the maximization is performed using the procedure `optmum` in the software package GAUSS. Here we propose a Bayesian hierarchical model and devise MCMC algorithms for statistical inference instead.

For the prior distributions, we let

$$\beta \sim \text{normal}(\beta_0, \Gamma_{\beta 0}), \quad \sigma_\nu^2 \sim \text{inverse gamma}(\sigma_{\nu 0}^2, \gamma_{\nu 0}), \quad \lambda \sim \text{uniform}(w^{-1}, 1) \quad (2.6)$$

where w denotes the smallest eigenvalue of W so that the resulting spatial variance-covariance matrix is positive definite. We will specify the prior distribution for $\{\mu_t\}$ in Sections 2.1-2.4. Even though here conjugate priors are used for β, σ_ν^2, and $\{\mu_t\}$, the hyper parameters are chosen such that the priors are diffuse. We perform sensitivity analysis to ensure that the posterior distribution is not sensitive to these choices. We use a Gibbs sampler here for simulating from the posterior distributions. For brevity, we present the full conditional distributions for the Gibbs sampler while omitting the detailed derivation. The full conditional distribution of β is,

$$\beta | y, \{\mu_t\}, \sigma_\nu^2, \lambda \sim \text{normal}(\beta_n, \Gamma_{\beta n}), \quad (2.7)$$

where $\beta_n = \Gamma_{\beta n}\left(\Gamma_{\beta 0}^{-1}\beta_0 + \sigma_\nu^{-2}\sum_{t=1}^T X_t'(I_N - \lambda W')(I_N - \lambda W)(y_t - \mu_t)\right)$ and $\Gamma_{\beta n}^{-1} = \Gamma_{\beta 0}^{-1} + \sigma_\nu^{-2}\sum_{t=1}^T X_t'(I_N - \lambda W')(I_N - \lambda W)X_t$. The full conditional distribution of σ_ν^2 is,

$$\sigma_\nu^2 | y, \beta, \{\mu_t\}, \lambda \sim \text{inverse gamma}(\sigma_{\nu n}^2, \gamma_{\nu n}), \quad (2.8)$$

where $\sigma_{\nu n}^2 = \sigma_{\nu 0}^2 + NT/2$ and $\gamma_{\nu n} = \gamma_{\nu 0} + 2^{-1}\sum_{t=1}^T (y_t - X_t\beta - \mu_t)'(I_N - \lambda W')(I_N - \lambda W)(y_t - X_t\beta - \mu_t)$. The full conditional distribution of λ is not in closed-form and

thus we use a Metropolis-Hastings algorithm with a normal distribution as the proposal distribution. In the case $\lambda = 0$, the disturbance model does not account for spatial dependence and reduces to white noise. The same Gibbs sampler above can be applied, with omission of the Metropolis-Hastings algorithm for simulating λ. We will specify the corresponding posterior distribution of μ_t in Sections 2.1-2.4.

2.1 Homogeneous model

In the homogeneous model, the state effects are assumed to be state-homogeneous and time-homogeneous with $\mu_t = \mu 1_N$, where μ is an overall mean and 1_N denotes an N dimensional vector of 1's. We use the following prior distribution for μ,

$$\mu \sim \text{normal}(\mu_0, \gamma_{\mu 0}). \tag{2.9}$$

Thus the full conditional distribution of μ is,

$$\mu | y, \beta, \sigma_\nu^2, \lambda \sim \text{normal}(\mu_n, \gamma_{\mu n}), \tag{2.10}$$

where $\mu_n = \gamma_{\mu n}(\gamma_{\mu 0}^{-1}\mu_0 + \sigma_\nu^{-2}\sum_{t=1}^{T} 1'_N(I_N - \lambda W')(I_N - \lambda W)(y_t - X_t\beta))$ and $\gamma_{\mu n} = \left(\gamma_{\mu 0}^{-1} + \sigma_\nu^{-2}T1'_N(I_N - \lambda W')(I_N - \lambda W)1_N\right)^{-1}$. When $\lambda = 0$, the model assumes spatial independence, which we will refer to as "homogeneous independent model". Otherwise when $\lambda \neq 0$, we will refer to the model as "homogeneous spatial model".

2.2 Heterogeneous model

In the heterogeneous model, the state effects are assumed to be state-homogeneous but time-inhomogeneous with $\mu_t = \mu_t 1_N$, where μ_t is a mean across states at time t. We let $\mu = (\mu_1, \ldots, \mu_T)'$ denote the vector of these means over time. Also, the regression coefficients and the spatial autocorrelation coefficient are assumed to be state homogeneous and time-inhomogeneous with $\beta = (\beta'_1, \ldots, \beta'_T)'$ and $\lambda = (\lambda_1, \ldots, \lambda_T)'$, where $\beta_t = (\beta_{1t}, \beta_{2t})'$. Then the distribution for the data at time t is,

$$y_t|\beta_t, \mu_t, \sigma_\nu^2, \lambda_t \quad \sim \quad \text{normal}\left(X_t\beta_t + \mu_t, \sigma_\nu^2(I_N - \lambda_t W)^{-1}(I_N - \lambda_t W')^{-1}\right) \tag{2.11}$$

Thus the distribution for the data at all time points is,

$$y|\beta, \mu, \sigma_\nu^2, \lambda \quad \sim \quad \text{normal}\left(\tilde{X}\beta + \mu, \sigma_\nu^2(I_{NT} - V)^{-1}(I_{NT} - V')^{-1}\right), \tag{2.12}$$

where $V = \text{diag}\{\lambda_t W\}_{t=1}^T$ and $\tilde{X} = (I_T \otimes J_N)X$ with J_N a $N \times 2$ matrix of all 1's. Let $\tilde{X}_t = (l_t \otimes J_N)X_t$ where l_t denotes the tth row of I_T. Thus $\tilde{X} = [\tilde{X}'_1, \ldots, \tilde{X}'_T]'$. With the following prior distributions for β and σ_ν^2,

$$\beta \sim \text{normal}(\beta_0, \Gamma_{\beta 0}), \quad \sigma_\nu^2 \sim \text{inverse gamma}(\sigma_{\nu 0}^2, \gamma_{\nu 0}), \tag{2.13}$$

we obtain the full conditional distribution of β,

$$\beta | y, \mu, \sigma_\nu^2, \lambda \sim \text{normal}(\beta_n, \Gamma_{\beta n}), \tag{2.14}$$

where $\beta_n = \Gamma_{\beta n} \left(\Gamma_{\beta 0}^{-1} \beta_0 + \sigma_\nu^{-2} \sum_{t=1}^T \tilde{X}'_t (I_N - \lambda_t W')(I_N - \lambda_t W)(y_t - \mu_t) \right)$, $\Gamma_{\beta n}^{-1} = \Gamma_{\beta 0}^{-1} + \sigma_\nu^{-2} \sum_{t=1}^T \tilde{X}'_t (I_N - \lambda_t W')(I_N - \lambda_t W)\tilde{X}_t$, and the full conditional distribution of σ_ν^2,

$$\sigma_\nu^2 | y, \beta, \mu, \lambda \sim \text{inverse gamma}(\sigma_{\nu n}^2, \gamma_{\nu n}), \tag{2.15}$$

where $\sigma_{\nu n}^2 = \sigma_{\nu 0}^2 + NT/2$ and $\gamma_{\nu n} = \gamma_{\nu 0} + 2^{-1} \sum_{t=1}^T (y_t - \tilde{X}_t \beta - \mu_t)'(I_N - \lambda_t W')(I_N - \lambda_t W)(y_t - \tilde{X}_t \beta - \mu_t)$. We let $\lambda_t \sim \text{uniform}(w^{-1}, 1)$. We use the following prior distribution for μ,

$$\mu \sim \text{normal}(\mu_0, \Gamma_{\mu 0}). \tag{2.16}$$

Thus the full conditional distribution of μ is,

$$\mu | y, \beta, \sigma_\nu^2, \lambda \sim \text{normal}(\mu_n, \Gamma_{\mu n}), \tag{2.17}$$

where $\mu_n = \Gamma_{\mu n}(\Gamma_{\mu 0}^{-1} \mu_0 + \sigma_\nu^{-2} B)$, $\Gamma_{\mu n}^{-1} = \Gamma_{\mu 0}^{-1} + \sigma_\nu^{-2} A$, A is a $T \times T$ diagonal matrix with diagonal elements $A(t, t) = 1'_N (I_N - \lambda_t W')(I_N - \lambda_t W)1_N$ and B is a T dimensional vector with the tth element $B(t) = 1'_N (I_N - \lambda_t W')(I_N - \lambda_t W)(y_t - \tilde{X}_t \beta)$. When $\lambda = 0$, the model assumes spatial independence, which we will refer to as "heterogeneous independent model". Otherwise when $\lambda \neq 0$, we will refer to the model as "heterogeneous spatial model".

2.3 Fixed-effects model

In the fixed-effects model, the state effects are assumed to be time-homogeneous but state-inhomogeneous with $\mu_t = \tilde{\mu} = (\mu_1, \ldots, \mu_N)'$, where μ_i is a mean over time for state i. We use the following prior distribution for $\tilde{\mu}$,

$$\tilde{\mu} \sim \text{normal}(\mu_0, \Gamma_{\mu 0}). \tag{2.18}$$

Thus the full conditional distribution of $\tilde{\mu}$ is,

$$\tilde{\mu} | y, \beta, \sigma_\nu^2, \lambda \sim \text{normal}(\mu_n, \Gamma_{\mu n}), \tag{2.19}$$

where $\mu_n = \Gamma_{\mu n}(\Gamma_{\mu 0}^{-1}\mu_0 + \sigma_\nu^{-2}\sum_{t=1}^{T}(I_N - \lambda W')(I_N - \lambda W)(y_t - X_t\beta))$ and $\Gamma_{\mu n}^{-1} = \Gamma_{\mu 0}^{-1} + \sigma_\nu^{-2}T(I_N - \lambda W')(I_N - \lambda W)$. When $\lambda = 0$, the model assumes spatial independence, which we will refer to as "fixed-effects independent model". Otherwise when $\lambda \neq 0$, we will refer to the model as "fixed-effects spatial model".

2.4 Random-effects model

In the random-effects model, the state effects are assumed to be time-homogeneous but state-inhomogeneous with $\mu_t = \tilde{\mu} = (\mu_1, \ldots, \mu_N)'$, where μ_i's are random variables following iid normal(μ^*, σ_μ^2). Then the full conditional distribution of $\tilde{\mu}$ is,

$$\tilde{\mu}|y, \beta, \sigma_\nu^2, \lambda, \mu^*, \sigma_\mu^2 \sim \text{normal}(\mu_n, \Gamma_{\mu n}), \tag{2.20}$$

where $\mu_n = \Gamma_{\mu n}(\sigma_\mu^{-2}\mu^* 1_N + \sigma_\nu^{-2}\sum_{t=1}^{T}(I_N - \lambda W')(I_N - \lambda W)(y_t - X_t\beta))$ and $\Gamma_{\mu n}^{-1} = \sigma_\mu^{-2}I_N + \sigma_\nu^{-2}T(I_N - \lambda W')(I_N - \lambda W)$. We use the following prior distribution for μ^* and σ_μ^2,

$$\mu^* \sim \text{normal}(\mu_0^*, \gamma_{\mu 0}^*), \quad \sigma_\mu^2 \sim \text{inverse gamma}(\sigma_{\mu 0}^2, \gamma_{\mu 0}). \tag{2.21}$$

The full conditional distribution of μ^* is,

$$\mu^*|\tilde{\mu}, \sigma_\mu^2 \sim \text{normal}(\mu_n^*, \gamma_{\mu n}^*), \tag{2.22}$$

where $\mu_n^* = \gamma_{\mu n}^*((\gamma_{\mu 0}^*)^{-1}\mu_0^* + \sigma_\mu^{-2}1_N'\tilde{\mu})$ and $\gamma_{\mu n}^* = ((\gamma_{\mu 0}^*)^{-1} + N\sigma_\mu^{-2})^{-1}$. The full conditional distribution of σ_μ^2 is,

$$\sigma_\mu^2|\tilde{\mu}, \mu^* \sim \text{inverse gamma}\left(\sigma_{\mu 0}^2 + N/2, \gamma_{\mu 0} + 2^{-1}(\tilde{\mu} - \mu^* 1_N)'(\tilde{\mu} - \mu^* 1_N)\right). \tag{2.23}$$

When $\lambda = 0$, the model assumes spatial independence, which we will refer to as "random-effects independent model". Otherwise when $\lambda \neq 0$, we will refer to the model as "random-effects spatial model".

3. New Spatial Panel Models and Bayesian Inference

The existing spatial panel models assume temporal independence of cigarette sales y_t, which may not be appropriate. To account for potential temporal dependence, we propose a spatial-temporal statistical model. Using the same notation as in Section 2, we consider the same data model as (2.1) and the same model for the state effect and disturbance as (2.2). But now, to account for temporal dependence, we let ϕ_t follow a spatial-temporal dependence model,

$$\phi_t = \lambda W \phi_t + \eta \phi_{t-1} + \nu_t, \tag{3.1}$$

where the new parameter η denotes a temporal autocorrelation coefficient and, in addition, ν_t is white noise iid $N(0, \sigma_\nu^2)$ and is independent of ϕ_{t-1}. In Sections 3.1-3.4, we will consider different specifications of μ_t that give rise to four new spatial panel models, corresponding to the homogeneous model, heterogeneous model, fixed-effects model, and random-effects model in Section 2.

Due to (3.1), we have the following distribution for the data at time t,

$$y_t | y_{t-1}, \beta, \mu_t, \mu_{t-1}, \sigma_\nu^2, \lambda, \eta \quad \sim \quad \text{normal} \left(\mu_{y_t}, \Sigma_{y_t} \right), \tag{3.2}$$

where $\mu_{y_t} = X_t \beta + \mu_t + \eta(I_N - \lambda W)^{-1}(y_{t-1} - X_{t-1}\beta - \mu_{t-1})$ and $\Sigma_{y_t} = \sigma_\nu^2 (I_N - \lambda W)^{-1}(I_N - \lambda W')^{-1}$. The spatial-temporal model may be viewed as a temporal process with the transition probability (3.2). When $\eta = 0$, model (3.1) reduces to (2.3). Thus the distribution for the data at all time points is,

$$y | \beta, \mu, \sigma_\nu^2, \lambda, \eta \quad \sim \quad \text{normal} \left(X\beta + \mu, \sigma_\nu^2 (I_{NT} - V)^{-1}(I_{NT} - V')^{-1} \right), \tag{3.3}$$

where $\mu = (\mu_1', \ldots, \mu_T')'$ and

$$V \quad = \quad \begin{pmatrix} \lambda W & 0 & 0 & \ldots \\ \eta I_N & \lambda W & 0 & \ldots \\ 0 & \eta I_N & \lambda W & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{3.4}$$

Then as in Section 2, we propose a Bayesian hierarchical model and devise MCMC algorithms for statistical inference.

We use the same prior distributions as in Section 2 for β, σ_ν^2 and λ (see (2.6)) and a uniform prior for η,

$$\eta \sim \text{uniform}(-1, 1). \tag{3.5}$$

We will specify the prior distribution for $\{\mu_t\}$ in Sections 3.1-3.4. Again we choose hyper parameters that ensure a diffuse prior for β, σ_ν^2 and $\{\mu_t\}$. We also perform sensitivity analysis to ensure that the posterior distribution is not sensitive to these choices. We use a Gibbs sampler here for simulating from the posterior distributions. Again for brevity we omit presenting the derivation of the full conditional distributions. The full conditional distribution of β is,

$$\beta | y, \{\mu_t\}, \sigma_\nu^2, \lambda, \eta \sim \text{normal}(\beta_n, \Gamma_{\beta n}), \tag{3.6}$$

where $\beta_n = \Gamma_{\beta n} \left(\Gamma_{\beta 0}^{-1} \beta_0 + \sigma_\nu^{-2} X'(I_{NT} - V')(I_{NT} - V)(y - \mu) \right)$ and $\Gamma_{\beta n}^{-1} = \Gamma_{\beta 0}^{-1} + \sigma_\nu^{-2} X'(I_{NT} - V')(I_{NT} - V)X$. The full conditional distribution of σ_ν^2 is,

$$\sigma_\nu^2 | y, \beta, \{\mu_t\}, \lambda, \eta \sim \text{inverse gamma}(\sigma_{\nu n}^2, \gamma_{\nu n}), \tag{3.7}$$

where $\sigma_{\nu n}^2 = \sigma_{\nu 0}^2 + NT/2$ and $\gamma_{\nu n} = \gamma_{\nu 0} + 2^{-1}(y - X\beta - \mu)'(I_{NT} - V')(I_{NT} - V)(y - X\beta - \mu)$. We use a Metropolis-Hastings algorithm with a normal distribution as the proposal distribution to update λ and η. In the case $\lambda = 0$, the disturbance model does not account for spatial dependence, but allows for temporal dependence. The corresponding posterior distribution of μ_t will be specified in Sections 3.1-3.4 using notation that is consistent with that in Section 2.

3.1 Homogeneous model

In the homogeneous model, the state effects are $\mu_t = \mu 1_N$ with an overall mean μ. We use the same prior distribution as (2.9) for μ. Thus the full conditional distribution of μ is,

$$\mu|y, \beta, \sigma_\nu^2, \lambda, \eta \sim \text{normal}(\mu_n, \gamma_{\mu n}), \tag{3.8}$$

where $\mu_n = \gamma_{\mu n}(\gamma_{\mu 0}^{-1}\mu_0 + \sigma_\nu^{-2}1'_{NT}(I_{NT} - V')(I_{NT} - V)(y - X\beta))$ and $\gamma_{\mu n} = (\gamma_{\mu 0}^{-1} + \sigma_\nu^{-2}1'_{NT}(I_{NT} - V')(I_{NT} - V)1_{NT})^{-1}$. When $\lambda = 0$, the model assumes spatial independence, which we will refer to as "homogeneous temporal model". Otherwise when $\lambda \neq 0$, we will refer to the model as "homogeneous spatial-temporal model".

3.2 Heterogeneous model

In the heterogeneous model, the state effects are $\mu_t = \mu_t 1_N$ with a mean across states μ_t at time t. We let $\mu = (\mu_1, \ldots, \mu_T)'$ denote the vector of these means over time. Also, the regression coefficients, spatial autocorrelation coefficient and temporal autocorrelation coefficient are assumed to be state homogeneous but time-inhomogeneous with $\beta = (\beta_1', \ldots, \beta_T')'$, $\lambda = (\lambda_1, \ldots, \lambda_T)'$ and $\eta = (\eta_1, \ldots, \eta_{T-1})'$, where $\beta_t = (\beta_{1t}, \beta_{2t})'$. Then the distribution for the data at time t is,

$$y_t|y_{t-1}, \beta, \mu, \sigma_\nu^2, \lambda, \eta \quad \sim \quad \text{normal}(\mu_{y_t}, \Sigma_{y_t}), \tag{3.9}$$

where $\mu_{y_t} = \tilde{X}_t\beta + \mu_t + \eta_{t-1}(I_N - \lambda_t W)^{-1}(y_{t-1} - \tilde{X}_{t-1}\beta - \mu_{t-1})$ and $\Sigma_{y_t} = \sigma_\nu^2(I_N - \lambda_t W)^{-1}(I_N - \lambda_t W')^{-1}$ with $\tilde{X} = (I_T \otimes J_N)X$. Thus the distribution for the data at all time points is,

$$y|\beta, \mu, \sigma_\nu^2, \lambda, \eta \quad \sim \quad \text{normal}\left(\tilde{X}\beta + \mu, \sigma_\nu^2(I_{NT} - V)^{-1}(I_{NT} - V')^{-1}\right), \tag{3.10}$$

where

$$V = \begin{pmatrix} \lambda_1 W & 0 & 0 & \cdots \\ \eta_1 I_N & \lambda_2 W & 0 & \cdots \\ 0 & \eta_2 I_N & \lambda_3 W & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

For the same prior distributions of β and σ_ν^2 as (2.13), we obtain the full conditional distribution of β,

$$\beta | y, \mu, \sigma_\nu^2, \lambda, \eta \sim \text{normal}(\beta_n, \Gamma_{\beta n}), \qquad (3.11)$$

where $\beta_n = \Gamma_{\beta n} \left(\Gamma_{\beta 0}^{-1} \beta_0 + \sigma_\nu^{-2} \tilde{X}'(I_{NT} - V')(I_{NT} - V)(y - \mu) \right)$, $\Gamma_{\beta n}^{-1} = \Gamma_{\beta 0}^{-1} + \sigma_\nu^{-2} \tilde{X}'(I_{NT} - V')(I_{NT} - V)\tilde{X}$, and the full conditional distribution of σ_ν^2,

$$\sigma_\nu^2 | y, \beta, \mu, \lambda, \eta \sim \text{inverse gamma}(\sigma_{\nu n}^2, \gamma_{\nu n}), \qquad (3.12)$$

where $\sigma_{\nu n}^2 = \sigma_{\nu 0}^2 + NT/2$ and $\gamma_{\nu n} = \gamma_{\nu 0} + 2^{-1}(y - \tilde{X}\beta - \mu)'(I_{NT} - V')(I_{NT} - V)(y - \tilde{X}\beta - \mu)$. For the flat prior distribution of η, we obtain the full conditional distribution of η,

$$\eta | y, \beta, \{\mu_t\}, \sigma_\nu^2, \lambda \sim \text{normal}(\eta_n, \Gamma_{\eta n}), \qquad (3.13)$$

where η_n is a $T - 1$ dimensional vector with the tth element $\eta_n = ((y_t - X_t\beta - \mu_t)'(y_t - X_t\beta - \mu_t))^{-1}((y_t - X_t\beta - \mu_t)'(I_N - \lambda_{t+1}W)(y_{t+1} - X_{t+1}\beta - \mu_{t+1}))$, and $\Gamma_{\eta n}$ is a $(T - 1) \times (T - 1)$ diagonal matrix with diagonal elements $\Gamma_{\eta n}(t, t) = (y_t - X_t\beta - \mu_t)'(y_t - X_t\beta - \mu_t))^{-1}\sigma_\nu^2$. We use the same prior distribution as (2.16) for μ. Thus the full conditional distribution of μ is,

$$\mu | y, \beta, \sigma_\nu^2, \lambda, \eta \sim \text{normal}(\mu_n, \Gamma_{\mu n}), \qquad (3.14)$$

where $\mu_n = \Gamma_{\mu n}(\Gamma_{\mu 0}^{-1}\mu_0 + \sigma_\nu^{-2}(I_T \otimes 1_N)'(I_{NT} - V')(I_{NT} - V)(y - \tilde{X}\beta))$ and $\Gamma_{\mu n}^{-1} = \Gamma_{\mu 0}^{-1} + \sigma_\nu^{-2}(I_T \otimes 1_N)'(I_{NT} - V')(I_{NT} - V)(I_T \otimes 1_N)$. When $\lambda = 0$, the model assumes spatial independence, which we will refer to as "heterogeneous temporal model". Otherwise when $\lambda \neq 0$, we will refer to the model as "heterogeneous spatial-temporal model".

3.3 Fixed-effects model

In the fixed-effects model, the state effects $\mu_t = \tilde{\mu} = (\mu_1, \ldots, \mu_N)'$ with μ_i a mean over time for state i. We use the same prior distribution as (2.18) for $\tilde{\mu}$. Thus the full conditional distribution of $\tilde{\mu}$ is,

$$\tilde{\mu} | y, \beta, \sigma_\nu^2, \lambda, \eta \sim \text{normal}(\mu_n, \Gamma_{\mu n}), \qquad (3.15)$$

where $\mu_n = \Gamma_{\mu n}(\Gamma_{\mu 0}^{-1}\mu_0 + \sigma_\nu^{-2}(1_T \otimes I_N)'(I_{NT} - V')(I_{NT} - V)(y - X\beta))$ and $\Gamma_{\mu n}^{-1} = \Gamma_{\mu 0}^{-1} + \sigma_\nu^{-2}(1_T \otimes I_N)'(I_{NT} - V')(I_{NT} - V)(1_T \otimes I_N)$. When $\lambda = 0$, the model assumes spatial independence, which we will refer to as "fixed-effects temporal model". Otherwise when $\lambda \neq 0$, we will refer to the model as "fixed-effects spatial-temporal model".

3.4 Random-effects model

In the random-effects model, the state effects $\mu_t = \tilde{\mu} = (\mu_1, \ldots, \mu_N)'$ with μ_i's random variables following iid normal(μ^*, σ_μ^2). Then the full conditional distribution of $\tilde{\mu}$ is,

$$\tilde{\mu}|y, \beta, \sigma_\nu^2, \lambda, \eta, \mu^*, \sigma_\mu^2 \sim \text{normal}(\mu_n, \Gamma_{\mu n}), \qquad (3.16)$$

where $\mu_n = \Gamma_{\mu n}(\sigma_\mu^{-2}\mu^* 1_N + \sigma_\nu^{-2}(1_T \otimes I_N)'(I_{NT} - V')(I_{NT} - V)(y - X\beta))$ and $\Gamma_{\mu n}^{-1} = \sigma_\mu^{-2}I_N + \sigma_\nu^{-2}(1_T \otimes I_N)'(I_{NT} - V')(I_{NT} - V)(1_T \otimes I_N)$. The prior and posterior distributions for μ^* and σ_μ^2 are the same as (2.21)–(2.23) in Section 2.4. When $\lambda = 0$, the model assumes spatial independence, which we will refer to as "random-effects temporal model". Otherwise when $\lambda \neq 0$, we will refer to the model as "random-effects spatial-temporal model".

Table 1: The central 95% credible intervals and the medians for the parameters in (a) the independent models that ignore spatial and temporal dependence; (b) the spatial models that account for spatial dependence. The abbreviations are homo = homogeneous model, hetero = heterogeneous model, fixed = fixed-effects model, random = random-effects models. Also reported are the deviance information criterion (DIC) for each model fitting.

Model		(a) Independent Models				(b) Spatial Models			
parameter		homo	hetero	fixed	random	homo	hetero	fixed	random
price	2.5%	-0.718	-1.293	-0.515	-0.510	-1.005	-1.353	-0.847	-0.848
β_1	50%	-0.618	-1.172	-0.464	-0.460	-0.882	-1.236	-0.751	-0.747
	97.5%	-0.532	-1.058	-0.414	-0.400	-0.763	-1.105	-0.659	-0.646
income	2.5%	0.061	0.034	-0.291	-0.289	0.207	0.464	-0.198	-0.191
β_2	50%	0.112	0.420	-0.252	-0.250	0.287	0.533	-0.105	-0.106
	97.5%	0.167	0.484	-0.212	-0.204	0.377	0.600	0.009	0.008
spatial	2.5%	-	-	-	-	0.340	0.157	0.538	0.526
λ	50%	-	-	-	-	0.410	0.244	0.614	0.611
	97.5%	-	-	-	-	0.487	0.344	0.687	0.693
	DIC	-547.0	-846.2	-2219.9	-1754.0	-665.0	-864.5	-2473.0	-1984.6

4. Analysis of Spatial Panel of Cigarette Demand

4.1 The existing spatial panel models

Table 1 summarizes the 2.5%, 50%, 97.5% percentiles of the posterior distributions of the parameters in the existing spatial panel models. For the heterogeneous models, the time-inhomogeneous parameter estimates averaged over time

are shown as in Table 3 of Baltagi and Li (2004). For the four spatial models, details of implementation of the Metropolis within Gibbs algorithm are as follows. For the normal proposal distributions, we tune the standard deviations to reach an average Hastings ratio between 0.2 and 0.7 as recommended in Gelman *et al.* (2004). The total number of iterations in the Metropolis within Gibbs run is 500,000 with a burn-in length of 20,000. Then every $1,000^{th}$ of the Monto Carlo samples are used to form a Monto Carlo sample of size 500.

The primary model parameters of interest are the elasticities of cigarette price and income. In the case of price elasticity, the 95% credible intervals based on the posterior distributions of the existing spatial panel models all suggest a significant negative effect of cigarette price on cigarette sales. Specifically, for the independent models, the homogeneous independent model gives a posterior median price elasticity of -0.618, whereas the heterogeneous independent model yields an average posterior median price elasticity of -1.172. The posterior distributions of price elasticity in the fixed-effects independent model and the random-effects independent model are similar with posterior medians of -0.464 and -0.460 respectively, which are smaller than those in the homogeneous and heterogeneous independent models. For the corresponding four spatial models, the posterior median price elasticities are -0.882, -1.236, -0.751 and -0.747, respectively. We note that the posterior median price elasticities are larger in absolute values when spatial dependence is accounted for, which suggests that when spatial dependence is accounted for, cigarette price has a stronger effect on cigarette sales.

In the case of income elasticity, it is not quite clear what the relation is between cigarette sales and income. For the independent models, the posterior median income elasticity in the homogeneous independent model is 0.112 and the average posterior median income elasticity in the heterogeneous independent model is 0.420. There is a significant positive effect of income on cigarette sales based on the 95% credible intervals. However, the posterior median income elasticities in the fixed-effects independent model and the random-effects independent model are -0.252 and -0.250 respectively, indicating a negative relation between cigarette sales and income. Based on the 95% credible intervals, these negative price elasticities are also significant. When spatial autocorrelation is accounted for, the posterior median income elasticities become larger in the homogeneous spatial model and the heterogeneous spatial model with values 0.287 and 0.533 respectively. However, in the fixed-effects spatial model and the random-effects spatial model, the income elasticities are no longer significant, based on the 95% credible intervals.

The results of the spatial autocorrelation coefficient suggest strong positive spatial dependence in the disturbance. The posterior median spatial autocorrelation coefficient in the homogeneous spatial model is 0.410. In the heterogeneous

spatial model, the average posterior median spatial autocorrelation coefficient is 0.244, which is somewhat smaller. In the fixed-effects spatial model and the random-effects spatial model, the posterior median spatial autocorrelation coefficients are 0.614 and 0.611 respectively, suggesting a stronger spatial dependence. Based on the 95% credible intervals, the spatial dependence is significant. The posterior medians of 0.614 and 0.611 are close to the MLEs of 0.61 and 0.65 in Baltagi and Li (2004), where statistical inference is via maximum likelihood.

4.2 The new spatial panel models

Table 2 summarizes the 2.5%, 50%, 97.5% percentiles of the posterior distributions of the new spatial panel models. For all the new models, details of the implementation of the Metropolis within Gibbs algorithm are similar to Section 4.1.

Table 2: The central 95% credible intervals and the medians for the parameters in (a) the temporal models that ignore spatial dependence; (b) the spatial-temporal models that account for spatial dependence and temporal dependence. The abbreviations are homo = homogeneous model, hetero = heterogeneous model, fixed = fixed-effects model, random = random-effects models. Also reported are the deviance information criterion (DIC) for each model fitting.

Model		(a) Temporal Models				(b) Spatial-Temporal Models			
parameter		homo	hetero	fixed	random	homo	hetero	fixed	random
price	2.5%	-0.460	-0.502	-0.370	-0.364	-0.451	-0.510	-0.373	-0.374
β_1	50%	-0.413	-0.429	-0.330	-0.325	-0.402	-0.430	-0.320	-0.336
	97.5%	-0.365	-0.350	-0.292	-0.280	-0.356	-0.344	-0.291	-0.298
income	2.5%	0.320	0.168	-0.016	-0.005	0.397	0.176	0.075	0.093
β_2	50%	0.377	0.270	0.041	0.052	0.448	0.265	0.137	0.159
	97.5%	0.439	0.360	0.109	0.109	0.506	0.363	0.1965	0.220
spatial	2.5%	-	-	-	-	0.047	0.002	0.061	0.059
λ	50% -	-	-	-	-	0.062	0.022	0.078	0.080
	97.5%	-	-	-	-	0.078	0.045	0.098	0.104
temporal	2.5%	0.995	0.941	0.994	0.995	0.971	0.941	0.967	0.961
η	50%	0.999	0.953	0.998	0.998	0.984	0.952	0.986	0.983
	97.5%	0.9999	0.964	0.999	0.999	0.998	0.963	0.999	0.998
	DIC	-3534.5	-3772.1	-4177.1	-3804.4	-3604.6	-3744.2	-4248.8	-4138.7

In the case of price elasticity, the 95% credible intervals based on the posterior distributions of the new spatial panel models all suggest a significant negative effect of cigarette price on cigarette sales. Specifically, for the temporal models without spatial dependence, the homogeneous temporal model gives a posterior median price elasticity of -0.413. The heterogeneous temporal model yields an

average posterior median price elasticity of -0.429, which is slightly larger than that in the homogeneous temporal model. The posterior distributions of price elasticity for the fixed-effects temporal model and the random-effects temporal model are similar with posterior medians of -0.330 and -0.325 respectively, which are smaller than those in the homogeneous and heterogeneous temporal models. For the spatial-temporal models accounting for both spatial and temporal dependence, the posterior median price elasticities are -0.402, -0.430,-0.320, and -0.336, respectively, which are very close to the posterior median price elasticities in the temporal models without spatial dependence. Furthermore, comparing the posterior median price elasticities in the new spatial panel models to those in the existing spatial panel models, we note that the posterior median price elasticities are smaller when the temporal dependence is accounted for. That is, when the temporal dependence is accounted for, cigarette price seems to have less effect on cigarette sales.

In the case of income elasticity, among the temporal models without spatial dependence, the posterior median income elasticity in the homogeneous temporal model is 0.377 and the average posterior median income elasticity in the heterogeneous temporal model is 0.270. There is a significant positive effect of income on cigarette sales based on the 95% credible intervals. However, in the fixed-effects temporal model and the random-effects temporal model, the income elasticities are not significant based on the 95% credible intervals. For the spatial-temporal models accounting for both spatial and temporal dependence, the posterior median income elasticities are 0.448, 0.265, 0.137, and 0.159, respectively. The 95% credible intervals based of these four models all suggest a significant positive effect of income on cigarette sales. Furthermore, comparing the posterior median income elasticities in the new spatial panel models to those in the existing spatial panel models, we note that the posterior median income elasticities in the homogeneous models and heterogeneous models all have positive signs. However, the posterior median income elasticities are negative in the fixed-effects independent model and the random-effects independent model, insignificant in the fixed-effects spatial model, the random-effects spatial model, the fixed-effects temporal model and the random-effects temporal model, and positive in the fixed-effects spatial-temporal model and the random-effects spatial-temporal model.

As for spatial dependence, the posterior median spatial autocorrelation coefficients in the homogeneous spatial-temporal model, fixed-effects spatial-temporal model and random-effects spatial-temporal model are 0.062, 0.078 and 0.080, respectively. The average posterior median spatial coefficient in the heterogeneous spatial-temporal model is 0.022. Based on the 95% credible intervals, the spatial dependence is significant in all four models. Comparing the posterior median spatial autocorrelation coefficients in the new spatial panel models to those in

the existing spatial panel models, we note that the spatial dependence is much weaker when the temporal dependence is accounted for. On the other hand, the results of the temporal autocorrelation coefficient suggest a very strong positive temporal dependence in the disturbance. The posterior median temporal autocorrelation coefficients in all the new spatial panel models are very similar. Based on the 95% credible intervals, the temporal dependence is significant. Given the persistent nature of cigarette consumption behavior, it is important to take temporal dependence into account. Indeed, our results reveal that it appears more important to incorporate the temporal dependence than the spatial dependence.

5. Model Comparisons

5.1 Model selection based on DIC

Here we use deviance information criterion (DIC) to compare and select models (Spiegelhalter *et al.* (2002). Let θ denote the parameters in a given model. Then the DIC is given by:

$$\text{DIC} = \bar{D}(y) + p_D,$$

where $\bar{D}(y)$ is the average deviance $D(y, \theta)$ averaged over the range of possible parameter values (here all MCMC samples of θ from the posterior distribution), and $D(y, \theta)$ is the Bayesian deviance defined as

$$D(y, \theta) = -2 \log p(y|\theta).$$

where $p(y|\theta)$ is the likelihood function. The average deviance $\bar{D}(y)$ can be interpreted as a Bayesian measure of model fitting. The term p_D is the effective number of parameters defined as

$$p_D = \bar{D}(y) - D(y, \bar{\theta}),$$

where $\bar{\theta}$ is the average of MCMC samples of θ. The term p_D can be interpreted as a measure of model complexity. A smaller value of $\bar{D}(y)$ indicates a better model fit, while a smaller value of p_D indicates a more parsimonious model. With the two terms together, a smaller value of DIC indicates a better model.

Table 1 provides the DIC values for the existing spatial panel models. Among the independent models, the homogeneous independent model has the largest DIC value and the heterogeneous independent model has the second largest DIC value. A substantial improvement occurs when state-wise heterogeneity is accounted for, because the fixed-effects independent model and the random-effects independent model give much smaller DIC values. Overall the fixed-effects model independent has the smallest DIC value and is thus the best among the independent models. Similar statements can be made about the DIC values of the spatial models.
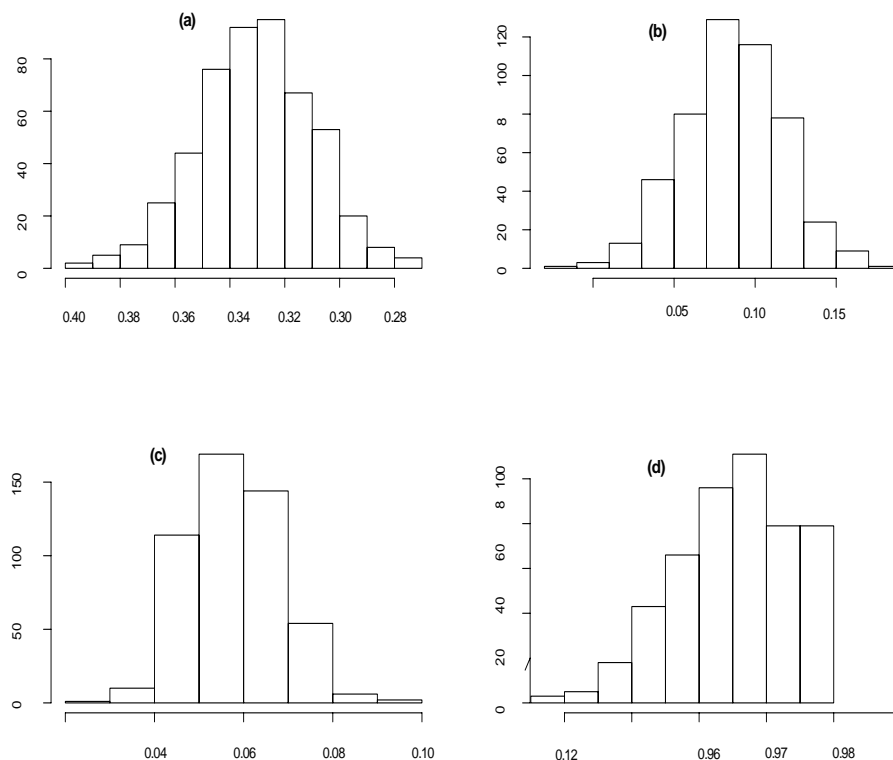
Figure 2: Posterior distribution of (a) cigarette price elasticity β_1; (b) income elasticity β_2; (c) spatial autocorrelation coefficient λ; (d) temporal autocorrelation coefficient η in the spatial-temporal fixed-effects model.

Furthermore, comparing the DIC values to those of independent models, we note that the models have somewhat lower DIC values when spatial dependence is accounted for.

Table 2 provides the DIC values for the new spatial panel models. Among the temporal models without spatial dependence, the homogeneous temporal model has the largest DIC value and the heterogeneous temporal model has the second largest DIC value. Accounting for the state-wise heterogeneity improves the models slightly. The fixed-effects temporal model and the random-effects temporal model give smaller DIC values. Overall the fixed-effects temporal model has the smallest DIC value and is thus the best among the temporal models. Similar observations are made about DIC values of the spatial-temporal models. Comparing Table 2 to Table 1, the DIC values of the new spatial panel models are much smaller than those of the existing spatial panel models, which suggests a substantial improvement of model fitting when temporal dependence is accounted for.

We show in Figure 2 the posterior distributions of the regression coefficients,

the spatial autocorrelation coefficient, and the temporal autocorrelation coefficient for the fixed-effects spatial-temporal model, which is the best among all the models. These posterior distributions show a significant negative relation between cigarette sales and cigarette price with a posterior median price elasticity of -0.320, and a significant positive relation between cigarette sales and income with a posterior median income elasticity of 0.137. According to economic theory, if cigarette is considered normal, then a consumer with a higher income would tend to buy more, which gives a positive income elasticity. On the other hand, income is highly correlated with education. Since those with more education are less likely to smoke, the income elasticity may be negative. The net effect, according to this spatial panel of data, seems to be a positive income elasticity. The temporal dependence is very strong with a posterior median temporal autocorrelation coefficient of 0.986, while spatial dependence is relatively weak with a posterior median spatial autocorrelation coefficient of 0.078.

Table 3: Root mean squared error of prediction using the existing spatial panel models

Models	1988	1989	1990	1991	1992	1989–1992
homogeneous independent	0.1945	0.2020	0.2236	0.2223	0.2013	0.2091
homogeneous spatial	0.1864	0.1890	0.2074	0.2004	0.1770	0.1923
heterogeneous independent	0.1944	0.1918	0.2054	0.1941	0.1688	0.1913
heterogeneous spatial	0.1947	0.1909	0.2035	0.1914	0.1684	0.1901
fixed-effects independent	0.1165	0.1260	0.1616	0.1765	0.1701	0.1521
fixed-effects spatial	0.1055	0.1092	0.1402	0.1459	0.1475	0.1310
random-effects independent	0.1173	0.1272	0.1631	0.1781	0.1715	0.1534
random-effects spatial	0.1056	0.1094	0.1404	0.1462	0.1475	0.1311

5.2 Model validation based on prediction

For the existing spatial panel models assuming temporal independence, prediction for 1988–1992 is made based on posterior predictive distribution. Let $\tilde{Y} = (Y'_{T+1}, \ldots, Y'_{T+S})'$ denote cigarette sales at future time points $T + 1, \ldots, T + S$. The posterior predictive distribution is

$$p(\tilde{Y}|Y) = \int p(\tilde{Y}|\theta)p(\theta|Y)d\theta$$

where the model parameters are $\theta = (\beta, \mu, \sigma_\nu^2)$ for independent models and $\theta = (\beta, \mu, \lambda, \sigma_\nu^2)$ for spatial models. For each MCMC posterior sample of θ, we generate $\tilde{Y}$ from the corresponding $p(\tilde{Y}|\theta)$ evaluated at this θ.

For the new spatial panel models, prediction is made using the transition probabilities specified in (3.2). In particular, for each MCMC sample of θ and

the cigarette demand in the Tth year, a sample of predicted cigarette demand in 46 states and the $(T+1)^{th}$ year is drawn according to the distribution in (3.2). Then based on the $(T+1)^{th}$ year, a sample of predicted cigarette demand in the $(T+2)^{th}$ year is drawn. This procedure is applied for five consecutive years from 1988 to 1992. Note that for the heterogeneous models, an average parameter estimate averaged over time is used. In all models, an average predicted value averaged over all MCMC samples of θ is used as the final predicted value.

For a given year, cigarette sales in 46 sates are predicted and then the corresponding root mean squared error (RMSE) is computed to assess prediction accuracy. An RMSE is defined as the square root of the mean squared difference between the actual observations and the predicted values. Table 3 gives the RMSE for prediction of cigarette demand for each year from 1988 to 1992 along with the RMSE for all five years, based on the existing spatial panel models. We focus our discussion on the RMSEs for all five years. Among the independent models, the homogeneous independent model gives the largest RMSE of 0.2091. The average heterogeneous independent model has a slightly smaller RMSE of 0.1913. By accounting for the state-wise heterogeneity, the fixed-effects independent model and the random-effects independent model have even smaller RMSEs of 0.1521 and 0.1534 respectively. Similarly, among the spatial models, the fixed-effects spatial model and the random-effects spatial model give smaller RMSEs of 0.1310 and 0.1311 respectively, while the homogeneous spatial model gives the largest RMSE of 0.1923. The RMSE of the heterogeneous spatial model is 0.1901. Comparing the RMSEs in the spatial models to those in the independent models, it is clear that accounting for spatial autocorrelation improves the prediction by reducing the RMSE. Moreover, the model that accounts for state-wise heterogeneity and spatial dependence is the best, according to the performance of prediction. That is, the fixed-effects spatial model and the random-effects spatial model are the best models among the existing spatial panel models. These findings are consistent with those in Baltagi and Li (2004).

Table 4: Root mean squared error of prediction using the the new spatial panel models

Models	1988	1989	1990	1991	1992	1989–1992
homogeneous temporal	0.0535	0.0741	0.1130	0.01239	0.1265	0.1025
homogeneous spatial-temporal	0.0462	0.0496	0.0735	0.0683	0.0808	0.0651
heterogeneous temporal	0.0579	0.0843	0.1292	0.1470	0.1498	0.1193
heterogeneous spatial-temporal	0.0559	0.0784	0.1204	0.1351	0.1351	0.1098
fixed-effects temporal	0.0509	0.0672	0.1054	0.1228	0.1235	0.0985
fixed-effects spatial-temporal	0.0469	0.0503	0.0772	0.0799	0.0880	0.0705
random-effects temporal	0.0479	0.0580	0.0863	0.0919	0.0921	0.0775
random-effects spatial-temporal	0.0462	0.0492	0.0764	0.0784	0.0878	0.0696

Table 4 gives the RMSEs for the prediction of cigarette demand for each year from 1988 to 1992 along with the RMSE for all five years, based on the new spatial panel models. Among the temporal models, the RMSEs in the homogeneous temporal model, the fixed-effects temporal model, and the random-effects temporal model are 0.1025, 0.0985, and 0.0775, respectively, which are quite close in value. The heterogeneous temporal model gives a slightly larger RMSE of 0.1193. Again accounting for the spatial dependence yields smaller RMSEs. Among the spatial-temporal models, the RMSEs for the homogeneous spatial-temporal model, the fixed-effects spatial-temporal model, and the random-effects spatial-temporal model are 0.0651, 0.0705, and 0.0696, respectively, which are again quite close. The heterogeneous spatial-temporal model also gives a slightly larger RMSE of 0.1098. Overall, the homogeneous spatial-temporal model, the fixed-effects spatial-temporal model, and the random-effects spatial-temporal model perform well in predicting cigarette demand. Comparing Table 4 to Table 3, the RMSEs in the new spatial panel models are much smaller than those in the existing spatial panel models, which suggests that accounting for temporal dependence improves prediction.

6. Conclusion and Discussion

In this article, we considered a demand equation to examine the effect of price and income on cigarette demand, based on a spatial panel of 46 states of the US over a 30-year period. We generalized the existing spatial panel models in Baltagi and Li (2004) to account for not only spatial dependence but also temporal dependence. For statistical inference, we adopted a fully Bayesian approach and developed MCMC algorithms to obtain posterior distributions of model parameters and posterior predictive distributions of cigarette demand at future time points. The proposed method here overcomes the computational obstacles in other approaches for similar models. We showed that the analysis results are comparable between MLE in Baltagi and Li (2004) and our Bayesian inference, based on the existing spatial panel models that do not account for temporal dependence. Moreover our new spatial panel models provide significant improvement over the existing spatial panel models in terms of model fitting and out-of-sample predictive power. Based on the best fixed-effects spatial-temporal model, we found a negative price elasticity as in Baltagi and Li (2004), but a positive income elasticity contrary to Baltagi and Li (2004). This further suggests the importance of accounting for spatial-temporal dependence simultaneously. We also found that the temporal dependence of cigarette demand appears stronger than the spatial dependence. Finally the methodology proposed here is suitable for spatial panel data analysis in general and may be applied to a wider range of data sets in spatial econometrics. Since the demand equation is of primary interest and the auto-

correlation in the disturbance is of secondary interest, we have used a relatively simple spatial-temporal model for the disturbance. Our modeling framework can be extended to accommodate more complex spatial-temporal dependence structure. For example, (3.1) can be generalized from an autoregressive order of 1 to p to account for stronger time dependence. Similarly, the temporal term ηI_N in (3.4) can be extended to more general forms to feature spatial-temporal interactions. For future research, it would be worthwhile to develop more general classes of spatial-temporal models for the disturbance, while ensuring that the computation remains feasible.

Acknowledgment

References

Anselin, L. (2006). Spatial panel econometrics. In *The Econometrics of Panel Data* (Edited by L. Mátyás and P. Sevestre), 4th edition. Springer.

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

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data*, 3rd edition. Wiley.

Baltagi, B. H. and Levin, D. (1986). Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation and advertising reconsidered. *The Review of Economics and Statistics* **68**, 148-155.

Baltagi, B. H., Song, S. H., Jung, B. C. and Koh, W. (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* **140**, 97-130.

Chen, X. and Conley, T. G. (2001). A new semiparametric spatial model for panel time series, *Journal of Econometrics* **105**, 59-83.

Cressie, N. (1993). *Statistics for Spatial Data*, revised edition. Wiley.

Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. journal = *The Review of Economics and Statistics* **80**,, 549-560.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models, *International Regional Science Review* **26**, 244-268.

Gelman, A., Carlin, J. B., Stern, H. and Rubin, D. (2004). *Bayesian Data Analysis*, 2nd edition. Chapman, Boca Raton.

Hamilton, J. L. (1972). The demand for cigarettes: Advertising, the Health Scare, and the cigarette advertising ban, journal = *The Review of Economics and Statistics* **54**, 401-411.

Kapoor, M., Kelejian, H. H. and Prucha, I. R. (2006). Panel data models with spatially correlated error components, *Journal of Econometrics*, in press.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit (with discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583-639.

Waller, L., Carlin, B. Xia, H. and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates, *Journal of the American Statistical Association* **92**, 607-617.

Zhu, J. Huang, H.-C. and Wu, J. (2005). Modeling spatial-temporal binary data using Markov random fields, *Journal of Agricultural, Biological, and Environmental Statistics* **10**, 212-225.

Zhu, J., Rasmussen, J., Moeller, J., Aukema, B. H. and Raffa, K. F. (2008). Spatial-temporal modeling of forest gaps generated by colonization from below- and above-ground bark beetle species. *Journal of the American Statistical Association* **103**, 162-177.

Yanbing Zheng
Department of Statistics
University of Kentucky
817 Patterson Office Tower
Lexington, Kentucky 40506, USA
yanbing.zheng@uky.edu

Jun Zhu
Department of Statistics
University of Wisconsin, Madison
1300 Univwersity Avenue
Madison, WI 53706, USA
jzhu@stat.wisc.edu

Dong Li
Department of Economics
Kansas State University
327 Waters Hall
Manhattan, Kansas 66506, USA
dongli@ksu.edu